

SDSS Galaxy Clustering: Luminosity & Color Dependence and Stochasticity

Molly E.C. Swanson and Max Tegmark

*Department of Physics and MIT Kavli Institute, Massachusetts Institute of Technology,
77 Massachusetts Ave, Cambridge, Massachusetts 02139, USA**

Michael Blanton

*Center for Cosmology and Particle Physics, Dept. of Physics,
New York University, 4 Washington Pl., New York, NY 10003, USA*

Idit Zehavi

*Dept. of Astronomy, Case Western Reserve University,
10900 Euclid Avenue, Cleveland, OH 44106-7215*

(Dated: Submitted to Phys. Rev. D. February 22, 2007)

Differences in clustering properties between galaxy subpopulations complicate the cosmological interpretation of the galaxy power spectrum, but can also provide insights about the physics underlying galaxy formation. To study the nature of this relative clustering, we perform a counts-in-cells analysis of galaxies in the Sloan Digital Sky Survey (SDSS) in which we measure the relative bias between pairs of galaxy subsamples of different luminosities and colors. We use a generalized χ^2 test to determine if the relative bias between each pair of subsamples is consistent with the simplest deterministic linear bias model, and we also use a maximum likelihood technique to further understand the nature of the relative bias between each pair. We find that the simple, deterministic model is a good fit for the luminosity-dependent bias on scales above $\sim 5 h^{-1} \text{Mpc}$, which is good news for using magnitude-limited surveys for cosmology. However, the color-dependent bias shows evidence for stochasticity and/or nonlinearity which increases in strength toward smaller scales, in agreement with previous studies of stochastic bias. Also, confirming hints seen in earlier work, the luminosity-dependent bias for red galaxies is significantly different from that of blue galaxies: both luminous and dim red galaxies have higher bias than moderately bright red galaxies, whereas the biasing of blue galaxies is not strongly luminosity-dependent. These results can be used to constrain galaxy formation models and also to quantify how the color and luminosity selection of a galaxy survey can impact measurements of the cosmological matter power spectrum.

I. INTRODUCTION

In order to use galaxy surveys to study the large-scale distribution of matter, the relation between the galaxies and the underlying matter — known as the *galaxy bias* — must be understood. Developing a detailed understanding of this bias is important for two reasons: bias is a key systematic uncertainty in the inference of cosmological parameters from galaxy surveys, and it also has implications for galaxy formation theory.

Since it is difficult to measure the dark matter distribution directly, we can gain insight by studying *relative bias*, i.e., the relation between the spatial distributions of different galaxy subpopulations. There is a rich body of literature on this subject tracing back many decades (see, e.g., [1–10]), and been studied extensively in recent years as well, both theoretically [11–20] and observationally [21–66]. Such studies have established that biasing depends on the type of galaxy under consideration — for example, early-type, red galaxies are more clustered than late-type, blue galaxies [21–55], and luminous galaxies are more clustered than dim galaxies [32–66]. Since different types of galaxies do not exactly trace each other,

it is thus impossible for them all to be exact tracers of the underlying matter distribution.

More quantitatively, the luminosity dependence of bias has been measured in the 2 Degree Field Galaxy Redshift Survey (2dFGRS; [67]) [35, 57] and in the Sloan Digital Sky Survey (SDSS; [68, 69]) [42, 50, 58] as well as other surveys, and it is generally found that luminous galaxies are more strongly biased, with the difference becoming more pronounced above L_* , the characteristic luminosity of a galaxy in the Schechter luminosity function [70].

These most recent studies measured the bias from ratios of correlation functions or power spectra. The variances of clustering estimators like correlation functions and power spectra are well-known to be the sum of two physically separate contributions: Poisson shot noise (due to the sampling of the underlying continuous density field with a finite number of galaxies) and sample variance (due to the fact that only a finite spatial volume is probed). On the large scales most relevant to cosmological parameter studies, sample variance dominated the aforementioned 2dFGRS and SDSS measurements, and therefore dominated the error bars on the inferred bias.

This sample variance is easy to understand: if the power spectrum of distant luminous galaxies is measured to be different than that of nearby dim galaxies, then part of this measured bias could be due to the nearby region happening to be more/less clumpy than the dis-

*Electronic address: molly@space.mit.edu

tant one. In this paper, we will eliminate this annoying sample variance by comparing how different galaxies cluster in the *same* region of space, extending the counts-in-cells work of [23, 24, 43] and the correlation function work of [35, 42, 50, 57]. Here we use the counts-in-cells technique: we divide the survey volume into roughly cubical cells and compare the number of galaxies of each type within each cell. This yields a local, point-by-point measure of the relative bias rather than a global one as in the correlation function method. In other words, by comparing two galaxy density fields directly in real space, including the phase information that correlation function and power spectrum estimators discard, we are able to provide substantially sharper bias constraints.

This local approach also enables us to quantify so-called stochastic bias [71–74]. It is well-known that the relation between galaxies and dark matter or between two different types of galaxies is not necessarily deterministic — galaxy formation processes that depend on variables other than the local matter density give rise to stochastic bias as described in [71–74]. Evidence for stochasticity in the relative bias between early-type and late-type galaxies has been presented in [23, 24, 28, 43], and [75] finds evidence for stochastic bias between galaxies and dark matter via weak lensing. The time evolution of such stochastic bias has been modeled in [72] and was recently updated in [76]. Stochasticity is even predicted in the relative bias between virialized clumps of dark matter (halos) and the linearly-evolved dark matter distribution [77, 78]. Here we aim to test whether stochasticity is necessary for modeling the luminosity-dependent or the color-dependent relative bias.

In this paper, we study the relative bias as a function of scale using a simple stochastic biasing model by comparing pairs of SDSS galaxy subsamples in cells of varying size. Such a study is timely for two reasons. First of all, the galaxy power spectrum has recently been measured to high precision on large scales with the goal of constraining cosmology [58, 79–81]. As techniques continue to improve and survey volumes continue to grow, it is necessary to reduce systematic uncertainties in order to keep pace with shrinking statistical uncertainties. A detailed understanding of complications due to the dependence of galaxy bias on scale, luminosity, and color will be essential for making precise cosmological inferences with the next generation of galaxy redshift surveys [82–87].

Secondly, a great deal of theoretical progress on models of galaxy formation has been made in recent years, and 2dFGRS and SDSS contain a large enough sample of galaxies that we can now begin to place robust and detailed observational constraints on these models. The framework known as the halo model [88] (see [89] for a comprehensive review) provides the tools needed to make comparisons between theory and observations. The halo model assumes that all galaxies form in dark matter halos, so the galaxy distribution can be modeled by first determining the halo distribution — either analytically

[90–92] or using N -body simulations [17, 93–95] — and then populating the halos with galaxies. This second step can be done using semi-analytical galaxy formation models [96–102] or with a statistical approach using a model for the halo occupation distribution (HOD) [97, 103–105] or conditional luminosity function (CLF) [13, 106] which prescribes how galaxies populate halos.

Although there are some concerns that the halo model does not capture all of the relevant physics [107–111], it has been applied successfully in a number of different contexts [51, 99, 112–116]. The correlation between a galaxy’s environment (i.e., the local density of surrounding galaxies) and its color and luminosity [38, 41] has been interpreted in the context of the halo model [15, 48, 117–119], and [13, 16, 120] make predictions for the bias as a function of galaxy type and luminosity using the CLF formalism. Additionally, [42, 83, 119, 121] use correlation function methods to study the luminosity and color dependence of galaxy clustering, and interpret the results using the Halo Occupation Distribution (HOD) framework. The analysis presented here is complementary to this body of work in that the counts-in-cells method is sensitive to larger scales, uses a different set of assumptions, and compares the two density fields directly in each cell rather than comparing ratios of their second moments. The halo model provides a natural framework in which to interpret the luminosity and color dependence of galaxy biasing statistics we measure here.

The rest of this paper is organized as follows: Section II describes our galaxy data, and Sections III A and III B describe the construction of our galaxy samples and the partition of the survey volume into cells. In Section III C we outline our relative bias framework, and in Sections III D and III E we describe our two main analysis methods. We present our results in Section IV and conclude with a qualitative interpretation of our results in the halo model context in Section V.

II. SDSS GALAXY DATA

The SDSS [68, 69] uses a mosaic CCD camera [122] on a dedicated telescope [123] to image the sky in five photometric bandpasses denoted u , g , r , i and z [124]. After astrometric calibration [125], photometric data reduction [126], and photometric calibration [127–130], galaxies are selected for spectroscopic observations. To a good approximation, the main galaxy sample consists of all galaxies with r -band apparent Petrosian magnitude $r < 17.77$ after correction for reddening as per [131]; there are about 90 such galaxies per square degree, with a median redshift of 0.1 and a tail out to $z \sim 0.25$. Galaxy spectra are also measured for the Luminous Red Galaxy sample [132], which is not used in this paper. These targets are assigned to spectroscopic plates of diameter 2.98° by an adaptive tiling algorithm [133] and observed with a pair of CCD spectrographs [134], after which the spec-

troscopic data reduction and redshift determination are performed by automated pipelines. The rms galaxy redshift errors are of order 30 km/s for main galaxies, hence negligible for the purposes of the present paper.

Our analysis is based on 380,614 main galaxies (the “safe0” cut) from the 444,189 galaxies in the 5th SDSS data release (“DR5”) [135], processed via the SDSS data repository at New York University [136]. The details of how these samples were processed and modeled are given in Appendix A of [58] and in [137]. The bottom line is that each sample is completely specified by three entities:

1. The galaxy positions (RA, Dec, and comoving redshift space distance r for each galaxy)
2. The radial selection function $\bar{n}(r)$, which gives the expected (not observed) number density of galaxies as a function of distance
3. The angular selection function $\bar{n}(\hat{r})$, which gives the completeness as a function of direction in the sky, specified in a set of spherical polygons [138].

The three-dimensional selection functions of our samples are separable, i.e., simply the product $\bar{n}(\mathbf{r}) = \bar{n}(\hat{r}) \bar{n}(r)$ of an angular and a radial part; here $r \equiv |\mathbf{r}|$ and $\hat{r} \equiv \mathbf{r}/r$ are the radial comoving distance and the unit vector corresponding to the position \mathbf{r} . The volume-limited samples used this paper are constructed so that their radial selection function $\bar{n}(r)$ is constant over a range of r and zero elsewhere. The effective sky area covered is $\Omega \equiv \int \bar{n}(\hat{r}) d\Omega \approx 5183$ square degrees, and the typical completeness $\bar{n}(\hat{r})$ exceeds 90%. The conversion from redshift z to comoving distance was made for a flat Λ CDM cosmological model with $\Omega_m = 0.25$. Additionally, we make a first-order correction for redshift space distortions by applying the finger-of-god compression algorithm described in [58] with a threshold density of $\delta_c = 200$.

III. ANALYSIS METHODS

A. Overlapping Volume-Limited Samples

The basic technique used in this paper is pairwise comparison of the three-dimensional density fields of galaxy samples with different colors and luminosities. As in [42], we focus on these two properties (as opposed to morphological type, spectral type, or surface brightness) for two reasons: they are straightforward to measure from SDSS data, and recent work [41] has found that luminosity and color is the pair of properties that is most predictive of the local overdensity. Since color and spectral type are strongly correlated, our study of the color dependence of bias probes similar physics as studies using spectral type [23, 24, 28, 35, 43].

Our base sample of SDSS galaxies (“safe0”) has an r -band apparent magnitude range of $14.5 < r < 17.5$. Following the method used in [58], we created a series of

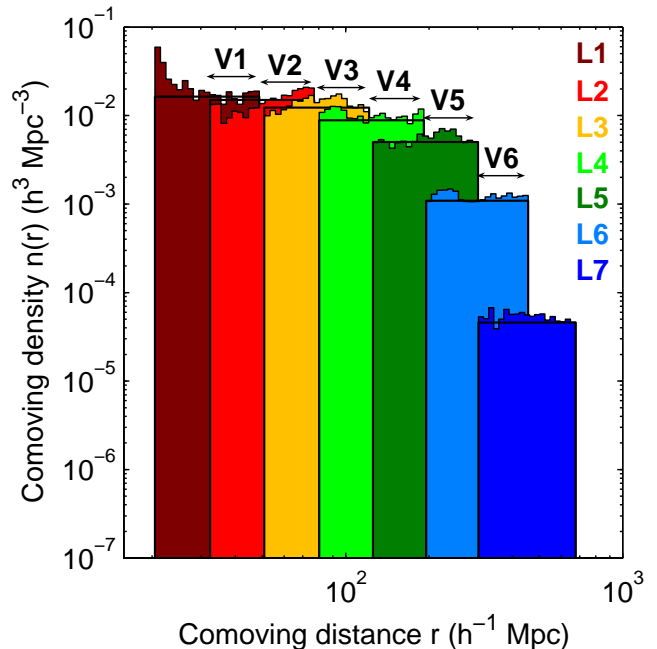


Figure 1: (Color online) Histogram of the comoving number density of the volume-limited samples L1-L7. The cuts used to define these samples are shown in Table I. Note that the radial selection function $\bar{n}(r)$ is uniform over the allowed range for each sample. Arrows indicate volumes V1-V6 where neighboring volume-limited samples overlap.

volume-limited samples containing galaxies in different luminosity ranges. These samples are defined by selecting a range of absolute magnitude $M_{\text{luminous}} < M_{0.1r} < M_{\text{dim}}$ and defining a redshift range such that the near limit has $M_{0.1r} = M_{\text{luminous}}$, $r = 14.5$ and the far limit has $M_{0.1r} = M_{\text{dim}}$, $r = 17.5$. Thus by discarding all galaxies outside the redshift range, we are left with a sample with a uniform radial selection function $\bar{n}(r)$ that contains all of the galaxies in the given absolute magnitude range in the volume defined by the redshift limits. Here $M_{0.1r}$ is defined as the absolute magnitude in the r -band shifted to a redshift of $z = 0.1$ [139].

Our volume-limited samples are labeled L1 through L7, with L1 being the dimmest and L7 being the most luminous. Figure 1 shows a histogram of the comoving galaxy density $n(r)$ for L1-L7. The cuts used to make these samples are shown in Table I.

Each sample overlaps spatially only with the samples in neighboring luminosity bins — since the apparent magnitude range spans three magnitudes and the absolute magnitude ranges for each bin span one magnitude, the far redshift limit of a given luminosity bin is approximately equal to the near redshift limit of the bin two notches more luminous. (It is not precisely equal due to evolution and K-corrections.)

The regions where neighboring volume-limited samples overlap provide a clean way to select data for studying the luminosity-dependent bias. By using only the

Table I: Summary of cuts used to create luminosity-binned volume-limited samples.

Luminosity-binned volume-limited samples	Absolute magnitude	Apparent magnitude	Redshift
L1	$-17 < M_{0.1r} < -16$	$14.5 < r < 17.5$	$0.007 < z < 0.016$
L2	$-18 < M_{0.1r} < -17$	$14.5 < r < 17.5$	$0.011 < z < 0.026$
L3	$-19 < M_{0.1r} < -18$	$14.5 < r < 17.5$	$0.017 < z < 0.041$
L4	$-20 < M_{0.1r} < -19$	$14.5 < r < 17.5$	$0.027 < z < 0.064$
L5	$-21 < M_{0.1r} < -20$	$14.5 < r < 17.5$	$0.042 < z < 0.100$
L6	$-22 < M_{0.1r} < -21$	$14.5 < r < 17.5$	$0.065 < z < 0.152$
L7	$-23 < M_{0.1r} < -22$	$14.5 < r < 17.5$	$0.101 < z < 0.226$

Table II: Overlapping volumes in which neighboring luminosity bins are compared.

Pairwise comparison (overlapping) volumes	Overlapping bins	Redshift
V1	L1 & L2	$0.011 < z < 0.016$
V2	L2 & L3	$0.017 < z < 0.026$
V3	L3 & L4	$0.027 < z < 0.041$
V4	L4 & L5	$0.042 < z < 0.064$
V5	L5 & L6	$0.065 < z < 0.100$
V6	L6 & L7	$0.101 < z < 0.152$

galaxies in the overlapping region from each of the two neighboring luminosity bins, we obtain two sets of objects (one from the dimmer bin and one from more luminous bin) whose selection is volume-limited and redshift-independent. Furthermore, since they occupy the same volume, they are correlated with the same underlying matter distribution, which eliminates uncertainty due to sample variance. We label the overlapping volume regions V1 through V6, where V1 is defined as the overlap between L1 and L2, and so forth. The redshift ranges for V1-V6 are shown Table II.

To study the color dependence of the bias, we further divide each sample into red galaxies and blue galaxies. Figure 2 shows the galaxy distribution of our volume-limited samples on a color-magnitude diagram. The sharp boundaries between the different horizontal slices are due to the differences in density and total volume sampled in each luminosity bin. This diagram illustrates the well-known color bimodality, with the redder galaxies falling predominantly in a region commonly known as the E-S0 ridgeline [140–142]. To separate the E-S0 ridgeline from the rest of the population, we use the same magnitude-dependent color cut as in [42]: we define galaxies with $^{0.1}(g-r) < 0.9 - 0.03(M_{0.1r} + 23)$ to be blue and galaxies on the other side of this line to be red.

In each volume V1-V6, we make four separate pairwise comparisons: luminous galaxies vs. dim galaxies, red galaxies vs. blue galaxies, luminous red galaxies vs. dim red galaxies, and luminous blue galaxies vs. dim blue galaxies. The luminous vs. dim comparisons measures the relative bias between galaxies in neighboring luminosity bins, and from this we can extract the luminosity dependence of the bias for all galaxies combined

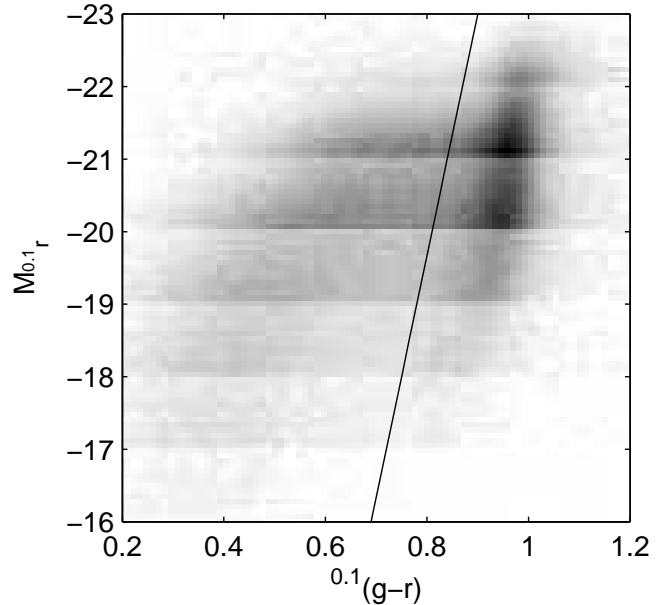


Figure 2: Color-magnitude diagram showing the number density distribution of the galaxies in the volume-limited samples. The shading scale has a square-root stretch, with darker areas indicating regions of higher density. The line shows the color cut of $^{0.1}(g-r) = 0.9 - 0.03(M_{0.1r} + 23)$. We refer to galaxies falling to the left of this line as blue and ones falling to the right of the line as red.

and for red and blue galaxies separately. The red vs. blue comparison measures the color-dependent bias. This set of four different types of pairwise comparisons is illustrated in Fig. 3 for V4, and the number of galaxies in each sample being compared is shown in Table III.

B. Counts-in-Cells Methodology

To compare the different pairs of galaxy samples, we perform a counts-in-cells analysis: we divide each comparison volume into roughly cubical cells and use the number of galaxies of each type in each cell as the primary input to our statistical analysis. This method is complementary to studies based on the correlation function since it involves point-by-point comparison of the

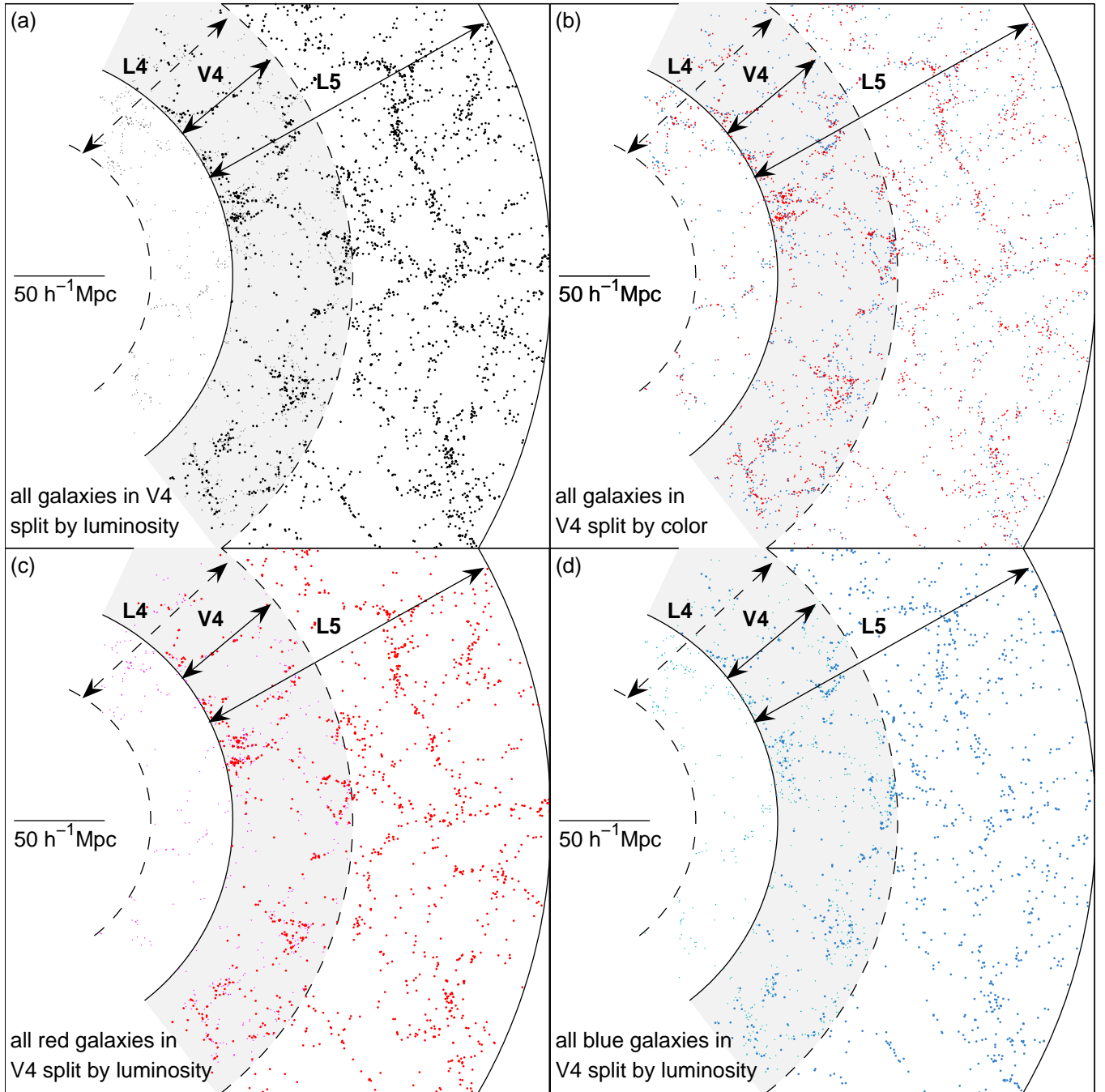


Figure 3: (Color online) Galaxy distributions plotted in comoving spatial coordinates for a radial slice of the volume-limited samples L4 (smaller dots, radial boundaries denoted by dashed lines) and L5 (larger dots, radial boundaries denoted by solid lines), which overlap in volume V4. Four different types of pairwise comparisons are illustrated: (a) luminous galaxies (L5) vs. dim galaxies (L4), (b) red galaxies vs. blue galaxies (both in V4), (c) luminous red galaxies (L5) vs. dim red galaxies (L4), and (d) luminous blue galaxies (L5) vs. dim blue galaxies (L4). The shaded regions denote the volume in which the two sets of galaxies are compared. A simple visual inspection shows that the different samples of galaxies being compared generally appear to cluster in the same physical locations — one key question we aim to answer here is if these observed correlations can be described with a simple linear bias model.

Table III: Number of galaxies in each sample being compared.

	All split by luminosity		All split by color	
	Luminous	Dim	Red	Blue
V1	427	651	125	953
V2	2102	2806	1117	3791
V3	6124	8273	5147	9250
V4	12122	23534	17144	18512
V5	11202	53410	37472	27140
V6	1784	38920	27138	13566
	Red split by luminosity		Blue split by luminosity	
	Red luminous	Red dim	Blue luminous	Blue dim
V1	72	53	355	598
V2	620	497	1482	2309
V3	2797	2350	3327	5923
V4	6848	10296	5274	13238
V5	7514	29958	3688	23452
V6	1451	25687	333	13233

two density fields and thus provides a more direct test of the local deterministic linear bias hypothesis. We probe scale dependence by varying the size of the cells.

To create our cells, we first divide the sky into two-dimensional “pixels” at four different angular resolutions using the SDSSPix pixelization scheme [143] as implemented by an updated version of the angular mask processing software MANGLE [138, 144]. The angular selection function $\bar{n}(\hat{r})$ is averaged over each pixel to obtain the completeness. To reduce the effects of pixels on the edge of the survey area or in regions affected by internal holes in the survey, we apply a cut on pixel completeness: we only use pixels with a completeness higher than 80% (50% for the lowest angular resolution). Figure 4 shows the pixelized SDSS angular mask at our four different resolutions, including only the pixels that pass our completeness cut. The different angular resolutions have 15, 33, 157, and 901 of these angular pixels respectively. At the lowest resolution, each pixel covers 353 square degrees, and the angular area of the pixels decreases by a factor of 1/4 at each resolution level, yielding pixels covering 88, 22, and 5 square degrees at the three higher resolutions.

To produce three-dimensional cells from our pixels, we divide each comparison volume into radial shells of equal volume. We choose the number of radial subdivisions at each angular resolution in each comparison volume such that our cells are approximately cubical, i.e., the radial extent of a cell is approximately equal to its transverse (angular) extent. This procedure makes cells that are not quite perfect cubes — there is some slight variation in the cell shapes, with cells on the near edge of the volume slightly elongated radially and cells on the far edge slightly flattened. We state all of our results as a function of cell size L , defined as the cube root of the cell volume. At the lowest resolution, there is just 1 radial shell for each volume; at the next resolution, we have 3 radial shells for volumes V4 and V5 and 2 radial shells for the other volumes. There are 5 radial shells at the

second highest resolution, and 10 at the highest.

Since each comparison volume is at a different distance from us, the angular geometry gives us cells of different physical size in each of the volumes. At the lowest resolution, where there is only one shell in each volume, the cell size is $14 h^{-1}\text{Mpc}$ in V1 and $134 h^{-1}\text{Mpc}$ in V6. At the highest resolution, the cell size is $1.7 h^{-1}\text{Mpc}$ in V1 and $16 h^{-1}\text{Mpc}$ in V6. Figure 5 shows the cells in each volume V1-V6 that are closest to a size of $\sim 20 h^{-1}\text{Mpc}$, the range in which the length scales probed by the different volumes overlap. (These are the cells used to produce the results shown in Fig. 8.)

C. Relative Bias Framework

Our task is to quantify the relationship between two fractional overdensity fields $\delta_1(\mathbf{x}) \equiv \rho_1(\mathbf{x})/\bar{\rho}_1 - 1$ and $\delta_2(\mathbf{x}) \equiv \rho_2(\mathbf{x})/\bar{\rho}_2 - 1$ representing two different types of objects. This framework is commonly used with types (1,2) representing (dark matter, galaxies), or as in [24, 28, 43], (early-type galaxies, late-type galaxies). Here we use it to represent (more luminous galaxies, dimmer galaxies) or (red galaxies, blue galaxies) to compare the samples described in Sec. III A. Galaxies are of course discrete objects, and as customary, we use the continuous field $\rho_\alpha(\mathbf{x})$ (where $\alpha=1$ or 2) to formally refer to the expectation value of the Poisson point process involved in distributing the type α galaxies.

The simplest (and frequently assumed) relationship between δ_1 and δ_2 is linear deterministic bias:

$$\delta_2(\mathbf{x}) = b_{\text{lin}} \delta_1(\mathbf{x}), \quad (1)$$

where b_{lin} is a constant parameter. This model cannot hold in all cases — note that it can give negative densities if $b_{\text{lin}} > 1$ — but is typically a reasonable approximation on cosmologically large length scales where the density fluctuations $\delta_1 \ll 1$, as is the case for the measurements of the large scale power spectrum recently used to constrain cosmological parameters [58, 79, 145]

More complicated models allow for nonlinearity and stochasticity, as described in detail in [73]:

$$\delta_2(\mathbf{x}) = b[\delta_1(\mathbf{x})] \delta_1(\mathbf{x}) + \epsilon(\mathbf{x}), \quad (2)$$

where the bias b is now a (typically slightly nonlinear) function of δ_1 . The stochasticity is represented by a random field ϵ — allowing for stochasticity removes the restriction implied by deterministic models that the peaks of δ_1 and δ_2 must coincide spatially. Stochasticity is basically the scatter in the relationship between the two density fields due to physical variables besides the local matter density. Nonlocal galaxy formation processes can also give rise to stochasticity, as discussed in [74].

We estimate the overdensity of galaxies of type α in cell i by

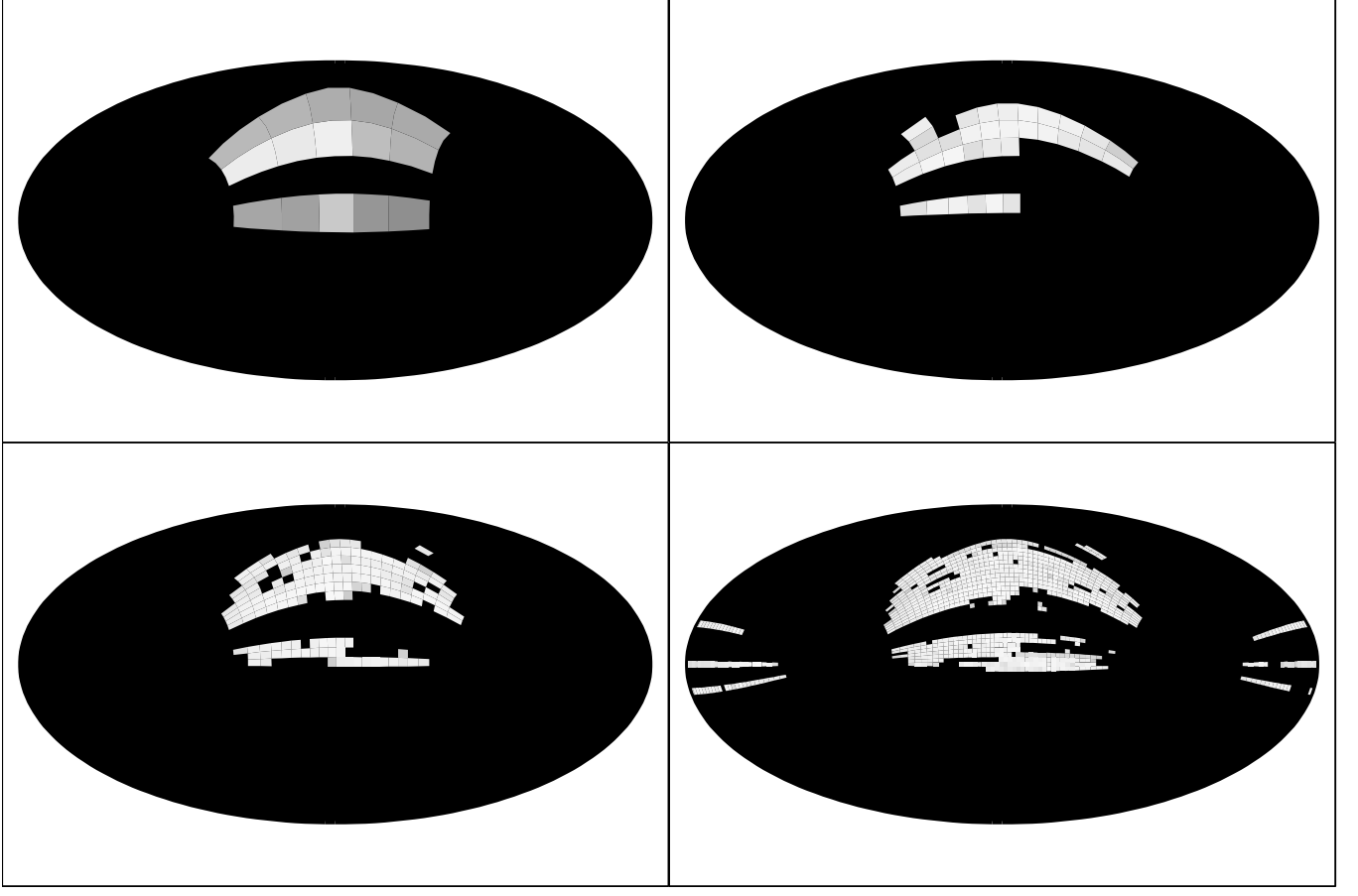


Figure 4: The SDSS DR5 angular mask pixelized at the four different resolutions used to partition the survey into cells, shown in Hammer-Aitoff projection in equatorial coordinates. Shading indicates completeness level: 0% is black, 100% is white.

$$g_{\alpha}^{(i)} \equiv \frac{N_{\alpha}^{(i)} - \bar{N}_{\alpha}^{(i)}}{\bar{N}_{\alpha}^{(i)}}, \quad (3)$$

where $N_{\alpha}^{(i)}$ is the number of observed type α galaxies in cell i and $\bar{N}_{\alpha}^{(i)}$ is the expected number of such galaxies, computed from the average angular selection function in the pixel and normalized so that the sum of $\bar{N}_{\alpha}^{(i)}$ over all cells in the comparison volume matches the total number of observed type α galaxies. The n -dimensional vectors

$$\mathbf{g}_{\alpha} \equiv \begin{pmatrix} g_{\alpha}^{(1)} \\ \vdots \\ g_{\alpha}^{(n)} \end{pmatrix} \quad (4)$$

contain the counts-in-cells data to which we apply the statistical analyses in Secs. IIID and IIIE.

The covariance matrix of \mathbf{g} is given by

$$\langle g_{\alpha}^{(i)} g_{\beta}^{(j)} \rangle = \langle \delta_{\alpha}^{(i)} \delta_{\beta}^{(j)} \rangle + \delta_{\alpha\beta} \mathbf{N}_{\alpha}^{ij}, \quad (5)$$

where $\delta_{\alpha}^{(i)}$ is the average of $\delta_{\alpha}(\mathbf{x})$ over cell i and, making

the customary assumption that the shot noise is Poissonian, the shot noise covariance matrix \mathbf{N}_{α} is given by

$$\mathbf{N}_{\alpha}^{ij} \equiv \delta_{ij} / \bar{N}_{\alpha}^{(i)}. \quad (6)$$

For comparing pairs of different types of galaxies, we construct the data vector

$$\mathbf{g} \equiv \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{pmatrix}, \quad (7)$$

which has a covariance matrix

$$\mathbf{C} \equiv \langle \mathbf{g} \mathbf{g}^T \rangle = \mathbf{S} + \mathbf{N}, \quad (8)$$

with

$$\mathbf{N} \equiv \begin{pmatrix} \mathbf{N}_1 & 0 \\ 0 & \mathbf{N}_2 \end{pmatrix}, \quad \mathbf{S} \equiv \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{12} & \mathbf{S}_{22} \end{pmatrix}, \quad (9)$$

and the elements of the matrix \mathbf{S} given by

$$\mathbf{S}_{\alpha\beta}^{ij} = \langle \delta_{\alpha}^{(i)} \delta_{\beta}^{(j)} \rangle. \quad (10)$$

Other counts-in-cells analyses often assume that the correlations between different cells can be ignored, i.e.,

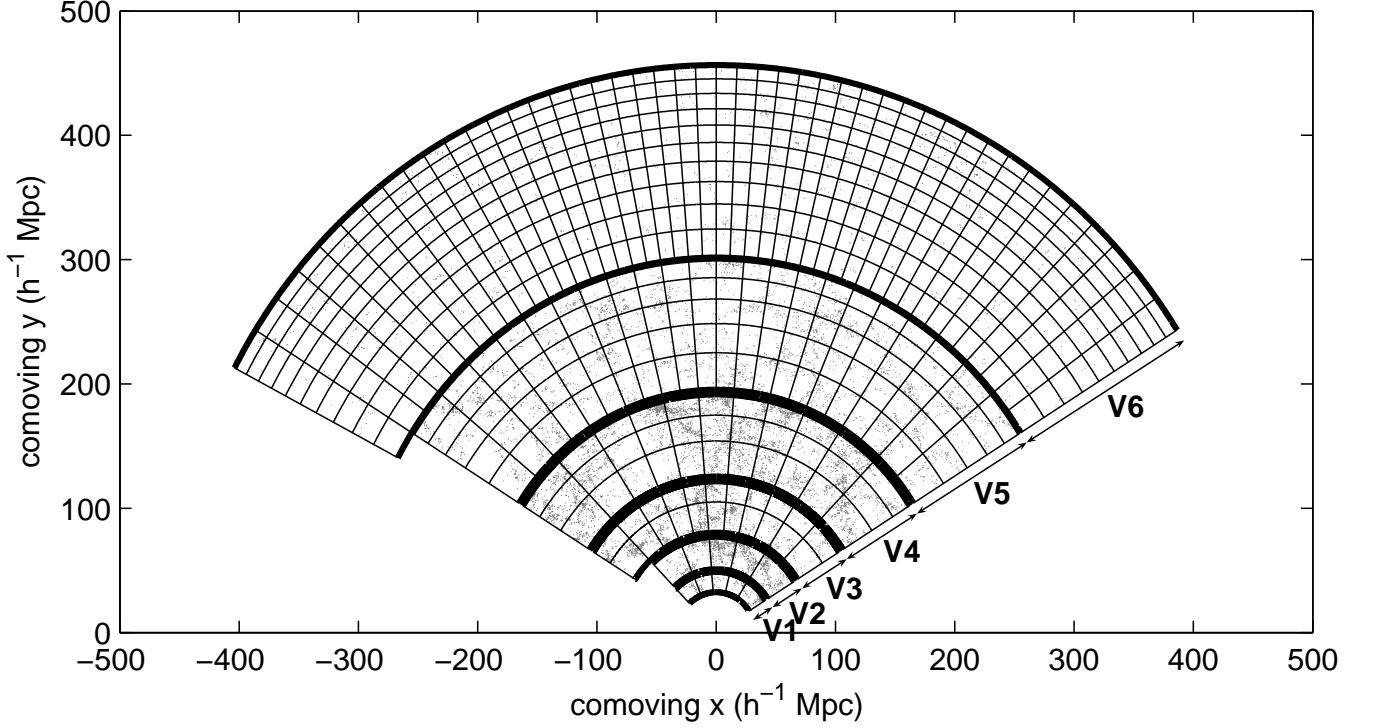


Figure 5: A radial slice of the SDSS survey volume divided into cells of size $\sim 20 h^{-1} \text{Mpc}$ with the galaxies in each cell shown in gray.

$\langle \delta_\alpha^{(i)} \delta_\beta^{(j)} \rangle = 0$ unless $i = j$. Here we account for cosmological correlations by computing the elements of \mathbf{S} using the best-fit ΛCDM matter power spectrum as we will now describe in detail. The power spectrum $P_{\alpha\beta}(\mathbf{k})$ is defined as $\langle \hat{\delta}_\alpha(\mathbf{k}) \hat{\delta}_\beta(\mathbf{k}')^\dagger \rangle = (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P_{\alpha\beta}(\mathbf{k})$, where $\hat{\delta}_\alpha(\mathbf{k}) \equiv \int e^{-i\mathbf{k}\cdot\mathbf{x}} \delta_\alpha(\mathbf{x}) d^3\mathbf{x}$ is the Fourier transform of the overdensity field. $P_{11}(\mathbf{k})$ and $P_{22}(\mathbf{k})$ are the power spectra of type 1 and 2 galaxies respectively, and $P_{12}(\mathbf{k})$ is the cross spectrum between type 1 and 2 galaxies. We assume isotropy and homogeneity, so that $P_{\alpha\beta}(\mathbf{k})$ is a function only of $k \equiv |\mathbf{k}|$, and rewrite the galaxy power spectra in terms of the matter power spectrum $P(k)$:

$$\begin{aligned} P_{11}(k) &= b_1(k)^2 P(k) \\ P_{12}(k) &= b_1(k) b_2(k) r_{12}(k) P(k) \\ P_{22}(k) &= b_2(k)^2 P(k), \end{aligned} \quad (11)$$

which defines the functions $b_1(k)$, $b_2(k)$, and $r_{12}(k)$.

To calculate $\langle \delta_\alpha^{(i)} \delta_\beta^{(j)} \rangle$ exactly, we need to convolve $\delta_\alpha(\mathbf{x})$ with a filter representing cell i and $\delta_\beta(\mathbf{x})$ with a filter representing cell j . This is complicated since our cells, while all roughly cubical, have slightly different shapes. We therefore use an approximation of a spherical top hat smoothing filter with radius R : $w(r, R) \equiv$

$3/(4\pi R^3) \Theta(R - r)$ with the Fourier transform given by

$$\hat{w}(k, R) = \frac{3}{(kR)^3} (\sin(kR) - kR \cos(kR)). \quad (12)$$

R is chosen so that the effective scale corresponds to cubes with side length L : $R = \sqrt{5/12} L$, where L is the cell size defined in Sec. III B. (See p. 500 in [146] for derivation of the $\sqrt{5/12}$ factor.) This gives

$$\langle \delta_\alpha^{(i)} \delta_\beta^{(j)} \rangle = \frac{1}{2\pi^2} \int_0^\infty \frac{\sin(kr_{ij})}{kr_{ij}} P_{\alpha\beta}(k) |\hat{w}(k, R)|^2 k^2 dk, \quad (13)$$

where r_{ij} is the distance between the centers of cells i and j . The kernel of this integrand — meaning everything besides $P_{\alpha\beta}(k)$ here — typically peaks at $k \sim 1/R$ and is only non-negligible in a range of $\Delta \log_{10} k \sim 1$. Assuming that the functions $b_1(k)$, $b_2(k)$, and $r_{12}(k)$ vary slowly with k over this range, they can be approximated by their values at $k_{\text{peak}} \equiv 1/R = \sqrt{12/5}/L$ and pulled outside the integral, allowing us to write

$$\mathbf{S} = \sigma_1^2(L) \begin{bmatrix} \mathbf{S}_M & b_{\text{rel}}(L) r_{\text{rel}}(L) \mathbf{S}_M \\ b_{\text{rel}}(L) r_{\text{rel}}(L) \mathbf{S}_M & b_{\text{rel}}(L)^2 \mathbf{S}_M \end{bmatrix} \quad (14)$$

where $\sigma_1^2(L) \equiv b_1(k_{\text{peak}})^2$, $b_{\text{rel}}(L) \equiv b_2(k_{\text{peak}})/b_1(k_{\text{peak}})$, $r_{\text{rel}}(L) \equiv r_{12}(k_{\text{peak}})$, and \mathbf{S}_M is the correlation matrix for the underlying matter density:

$$\mathbf{S}_M^{ij} = \frac{1}{2\pi^2} \int_0^\infty \frac{\sin(kr_{ij})}{kr_{ij}} P(k) |\hat{w}(k, R)|^2 k^2 dk. \quad (15)$$

For the matter power spectrum $P(k)$, we use the fitting formula from [147] with the best-fit “vanilla” parameters from [148] and apply the nonlinear transformation of [93].

Our primary parameters are the relative bias factor $b_{\text{rel}}(L)$, the relative cross-correlation coefficient $r_{\text{rel}}(L)$, and the overall normalization $\sigma_1^2(L)$. The only assumptions we have made in defining these parameters are homogeneity, isotropy, and that $b_1(k)$, $b_2(k)$, and $r_{12}(k)$ vary slowly in k . These parameters are closely related to those in the biasing models specified in equations (1) and (2): If linear deterministic biasing holds, then $b_{\text{rel}} = b_{\text{lin}}$ and $r_{\text{rel}} = 1$, and the addition of either nonlinearity or stochasticity will give $r_{\text{rel}} < 1$. As discussed in [74], stochasticity is expected to vanish in Fourier space (i.e., $r_{12}(k) = 1$) on large scales where the density fluctuations are small, but scale dependence of $b_1(k)$ and $b_2(k)$ can still give rise to stochasticity in real space. We will measure the parameters $b_{\text{rel}}(L)$ and $r_{\text{rel}}(L)$ as a function of scale, thus testing whether the bias is scale dependent and determining the range of scales on which linear biasing holds.

D. The Null-buster Test

Can the relative bias between dim and luminous galaxies or between red and blue galaxies be explained by simple linear deterministic biasing? To address this question, we use the so-called null-buster test described in [23]. For a pair of different types of galaxies, we calculate a difference map

$$\Delta \mathbf{g} \equiv \mathbf{g}_2 - f \mathbf{g}_1 \quad (16)$$

for a range of values of f . If equation (1) holds and $f = b_{\text{lin}}$, then the density fluctuations cancel and $\Delta \mathbf{g}$ will contain only shot noise, with a covariance matrix $\langle \Delta \mathbf{g} \Delta \mathbf{g}^T \rangle = \mathbf{N}_\Delta \equiv \mathbf{N}_1 + f^2 \mathbf{N}_2$ — this is our null hypothesis.

If equation (1) does not hold and the covariance matrix is instead given by $\langle \Delta \mathbf{g} \Delta \mathbf{g}^T \rangle = \mathbf{N}_\Delta + \mathbf{S}_\Delta$, where \mathbf{S}_Δ is some residual signal, then the most powerful test for ruling out the null hypothesis is the generalized χ^2 statistic [72]

$$\nu \equiv \frac{\Delta \mathbf{g}^T \mathbf{N}_\Delta^{-1} \mathbf{S}_\Delta \mathbf{N}_\Delta^{-1} \Delta \mathbf{g} - \text{Tr}(\mathbf{N}_\Delta^{-1} \mathbf{S}_\Delta)}{[2 \text{Tr}(\mathbf{N}_\Delta^{-1} \mathbf{S}_\Delta \mathbf{N}_\Delta^{-1} \mathbf{S}_\Delta)]^{1/2}}, \quad (17)$$

which can be interpreted as the significance level (i.e. the number of “sigmas”) at which we can rule out the null hypothesis. As detailed in [149], this test assumes that the Poissonian shot noise contribution can be approximated as Gaussian but makes no other assumptions about the probability distribution of $\Delta \mathbf{g}$. It is a valid

test for any choice of \mathbf{S}_Δ and reduces to a standard χ^2 test if $\mathbf{S}_\Delta = \mathbf{N}_\Delta$, but it rules out the null hypothesis with maximum significance in the case where \mathbf{S}_Δ is the true residual signal.

Using equations (9), (10), and (14), the covariance matrix of $\Delta \mathbf{g}$ can be written as

$$\langle \Delta \mathbf{g} \Delta \mathbf{g}^T \rangle = \sigma_1^2 (f^2 - 2b_{\text{rel}} r_{\text{rel}} f + b_{\text{rel}}^2) \mathbf{S}_M + \mathbf{N}_\Delta, \quad (18)$$

where \mathbf{S}_M is given by equation (15). We use $\mathbf{S}_\Delta = \mathbf{S}_M$ in equation (17) (note that ν is independent of the normalization of \mathbf{S} , which scales out) since deviations from linear deterministic bias are likely to be correlated with large-scale structure.

To apply the null-buster test, we compute ν as a function of f and then minimize it. If the minimum value $\nu_{\text{min}} > 2$, we rule out linear deterministic bias at $> 2\sigma$. If the null hypothesis cannot be ruled out and we are willing to assume that it is true, we can use the value of f that gives ν_{min} as a measure of b_{rel} . To calculate the uncertainty on b_{rel} , we use the range in f that gives $\sqrt{2N}(\nu - \nu_{\text{min}}) \leq 1$, where N is the number of degrees of freedom (equal to the number of cells minus 1 fitted parameter). This is a generalization of the standard $\Delta \chi^2 = 1$ uncertainty since ν is a generalization of $(\chi^2 - N)/\sqrt{2N}$.

E. Maximum Likelihood Method

In addition to the null-buster test, we use a maximum likelihood analysis to determine the parameters b_{rel} and r_{rel} . Our method is a generalization of the maximum likelihood method used in previous papers, accounting for correlations between different cells but making a somewhat different set of assumptions.

In [24, 28, 43], the probability of observing N_1 galaxies of type 1 and N_2 galaxies of type 2 in a given cell is expressed as

$$P(N_1, N_2) = \int_{-1}^\infty \int_{-1}^\infty \text{Pois} [N_1, \bar{N}_1 (1 + \delta_1)] \times \text{Pois} [N_2, \bar{N}_2 (1 + \delta_2)] f(\delta_1, \delta_2, \alpha) d\delta_1 d\delta_2, \quad (19)$$

where $f(\delta_1, \delta_2, \alpha)$ is the joint probability distribution of δ_1 and δ_2 in one cell, α represents a set of parameters which depend on the biasing model, and $\text{Pois}(N, \lambda) \equiv \lambda^N e^{-\lambda} / N!$ is the Poisson probability to observe N objects given a mean value λ . The likelihood function for n cells is then given by

$$\mathcal{L}(\alpha) \equiv \prod_{i=1}^n P(N_1^{(i)}, N_2^{(i)}), \quad (20)$$

which is minimized with respect to the parameters α . This treatment makes two assumptions: it neglects correlations between different cells and it assumes that the galaxy discreteness is Poissonian. These assumptions

greatly simplify the computation of \mathcal{L} , but are understood to be approximations to the true process. Cosmological correlations are known to exist on large scales, although their impact on counts-in-cells analyses has been argued to be small [28, 150]. Semi-analytical modeling [104, 151], N -body simulation [77, 94], and smoothed particle hydrodynamic simulation [97] investigations suggest that the probability distribution for galaxies/halos is sub-Poissonian in some regimes, and in fact non-Poissonian behavior is implied by observations as well [43, 106].

Dropping these two assumptions, we can write a more general expression for the likelihood function for n cells:

$$\begin{aligned} \mathcal{L}(\alpha, \beta) \equiv & P\left(N_1^{(1)}, \dots, N_1^{(n)}, N_2^{(1)}, \dots, N_2^{(n)}\right) = \\ & \int_{-1}^{\infty} \dots \int_{-1}^{\infty} \left[\prod_{i=1}^n P_g\left(N_1^{(i)}, \bar{N}_1^{(i)}\left(1 + \delta_1^{(i)}\right), \beta\right) \right] \\ & \times \left[\prod_{j=1}^n P_g\left(N_2^{(j)}, \bar{N}_2^{(j)}\left(1 + \delta_2^{(j)}\right), \beta\right) \right] \\ & \times f\left(\delta_1^{(1)}, \dots, \delta_1^{(n)}, \delta_2^{(1)}, \dots, \delta_2^{(n)}, \alpha\right) \\ & \times d\delta_1^{(1)} \dots d\delta_1^{(n)} d\delta_2^{(1)} \dots d\delta_2^{(n)}, \end{aligned} \quad (21)$$

where $\text{Poiss}(N, \lambda)$ has been replaced with a generic probability $P_g(N, \lambda, \beta)$ for the galaxy distribution parameterized by some parameters β and $f\left(\delta_1^{(1)}, \dots, \delta_1^{(n)}, \delta_2^{(1)}, \dots, \delta_2^{(n)}, \alpha\right)$ is a joint probability distribution relating δ_1 and δ_2 in all cells. In practice, this would be prohibitively difficult to calculate as it involves $2n$ integrations [152], and would require a reasonable parameterized form for $P_g(N, \lambda, \beta)$ as well as $f\left(\delta_1^{(1)}, \dots, \delta_1^{(n)}, \delta_2^{(1)}, \dots, \delta_2^{(n)}, \alpha\right)$.

In this paper, we take a simpler approach and approximate the probability distribution for our data vector \mathbf{g} to be Gaussian with the covariance matrix \mathbf{C} as defined by equations (8), (9) and (14), and use this to define our likelihood function in terms of the parameters σ_1^2 , b_{rel} , and r_{rel} :

$$\begin{aligned} \mathcal{L}(\sigma_1^2, b_{\text{rel}}, r_{\text{rel}}) \equiv & P\left(g_1^{(1)}, \dots, g_1^{(n)}, g_2^{(1)}, \dots, g_2^{(n)}\right) \\ & = \frac{1}{(2\pi)^n |\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{g}^\dagger \mathbf{C}^{-1} \mathbf{g}\right). \end{aligned} \quad (22)$$

Note that this includes the shot noise since $\mathbf{C} = \mathbf{S} + \mathbf{N}$, and is not precisely equivalent to assuming that P_g and f in equation (21) are Gaussian.

To determine the best fit values of our parameters for each pairwise comparison, we maximize $2 \ln \mathcal{L}(\sigma_1^2, b_{\text{rel}}, r_{\text{rel}})$ with respect to σ_1^2 , b_{rel} , and r_{rel} . Since our method of comparing pairs of galaxy samples primarily probes the relative biasing between the two types of galaxies, it is not particularly sensitive to σ_1^2 , which represents the bias of type 1 galaxies relative to

the dark matter power spectrum used in equation (15). Thus we marginalize over σ_1^2 and calculate the uncertainty on b_{rel} and r_{rel} using the $\Delta(2 \ln \mathcal{L}) = 1$ contour in the $b_{\text{rel}}\text{-}r_{\text{rel}}$ plane.

As an alternate method of calculating the uncertainties, we also generate Monte Carlo realizations of the cell counts $N_\alpha^{(i)}$ using the probability distribution in equation (21) and apply the same maximum likelihood method we used for the real data. For the Monte Carlo generation, we take P_g to be Poissonian and f to be Gaussian with covariance matrix \mathbf{S} :

$$\begin{aligned} f\left(\delta_1^{(1)}, \dots, \delta_1^{(n)}, \delta_2^{(1)}, \dots, \delta_2^{(n)}, \alpha = \sigma_1^2, b_{\text{rel}}, r_{\text{rel}}\right) \\ = \frac{1}{(2\pi)^n |\mathbf{S}|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\delta}^\dagger \mathbf{S}^{-1} \boldsymbol{\delta}\right), \end{aligned} \quad (23)$$

where \mathbf{S} is given by equation (14) and $\boldsymbol{\delta}$ is a vector of the overdensities defined in the same manner as equation (7). In other words, our method is equivalent to first simulating smooth density fields corresponding to the two kinds of galaxies with the appropriate power spectrum, relative bias and correlation coefficient r , then distributing simulated galaxies according to a Poisson process modulated by these smooth fields, and finally binning these galaxies into cells. Note that this model is similar to the bivariate Gaussian model discussed in [73] which implies linear stochastic biasing, except that it includes correlations between different cells. For each pairwise comparison, mock data sets are generated using the best-fit values of the parameters σ_1^2 , b_{rel} , and r_{rel} from the data. The standard deviations of the best-fit parameter values from the mock data sets then provide a measure of the uncertainties.

IV. RESULTS

A. Null-buster Results

To test the deterministic linear bias model, we apply the null-buster test described in Sec. IIID to the pairs of galaxy samples described in Sec. IIIA. For studying the luminosity-dependent bias, we use the galaxies in the more luminous bin as the type 1 galaxies and the dimmer bin as the type 2 galaxies for each pair of neighboring luminosity bins, and repeat this in each volume V1-V6. We do this for all galaxies and also for red and blue galaxies separately. For the color dependence, we use red galaxies as type 1 and blue galaxies as type 2, and again repeat this in each volume. To determine the scale dependence, we repeat all of these tests for four different values of the cell size L as described in Sec. IIIB.

1. Is the bias linear and deterministic?

The results are plotted in Fig. 6, which shows the minimum value of the null-buster test statistic ν_{min} vs. cell

size L . According to this test, deterministic linear biasing is in fact an excellent fit for the luminosity-dependent bias: nearly all ν_{\min} fall within $|\nu_{\min}| < 2$, indicating consistency with the null hypothesis at the 2σ level. (There are a few exceptions in the case of the red galaxies, the largest being $\nu_{\min} \sim 5$ for the smallest cell size in V3.) For color-dependent bias, however, deterministic linear biasing is ruled out quite strongly, especially at smaller scales.

The cases where the null hypothesis survives are quite noteworthy, since this implies that essentially all of the large clustering signal that is present in the data (and is visually apparent in Figure 3) is common to the two galaxy samples and can be subtracted out. For example, for the V5 luminosity split at the highest angular resolution ($L = 10 h^{-1}\text{Mpc}$), clustering signal is detected at 953σ in the faint sample ($\nu(f) \approx 953$ for $f = 0$) and at 255σ in the bright sample ($\nu(f) \approx 255$ for $f = \infty$), yet the weighted difference of the two maps is consistent with mere shot noise ($\nu(0.88) \approx -0.63$). This also shows that no luminosity-related systematic errors afflict the sample selection even at that low level.

2. Is the bias independent of scale?

For the luminosity-dependent bias, we thus use the value of f that gives ν_{\min} as a measure of b_{rel} , the relative bias between two neighboring luminosity bins. We find that although the value of b_{rel} depends on luminosity, it does not appear to depend strongly on scale, as can be seen in Fig. 7: in all plots the curves appear roughly horizontal. To test this “chi-by-eye” inference of scale independence quantitatively, we applied a simple χ^2 fit on the four data points in each volume using a one-parameter model: a horizontal line with a constant value of b_{rel} . We define this model to be a good fit if the goodness-of-fit value (the probability that a χ^2 as poor as the value calculated should occur by chance, as defined in [153]) exceeds 0.01.

We find that this model of no scale dependence is a good fit for all data sets plotted in Fig. 7(a), (c), and (d) with the exception of the red galaxies in V2, and even these give a good fit if the smallest-scale data point is dropped, implying scale independence above $\sim 5 h^{-1}\text{Mpc}$. We therefore find no evidence that the luminosity-dependent bias is scale-dependent above the scale of galaxy clusters. This implies that recent cosmological parameter analyses which use only measurements on scales $\gtrsim 60 h^{-1}\text{Mpc}$ (e.g., [79, 154, 155]) are probably justified in assuming scale independence of luminosity-dependent bias.

In contrast, applying this χ^2 fit to the color-dependent bias, Fig. 7(b), does show evidence for scale dependence, with the bias being slightly lower at smaller scales. When fitting with all four data points, the scale-independent model is ruled out for V1-V5. If we drop the smallest-scale data points and apply the fit with only three data

points, scale independence is still ruled out for V1 and V5. Applying the fit with only the two largest-scale data points gives a good fit in all six volumes. To summarize this in terms of physical scales: the color-dependent bias in V1-V5 is scale-independent above $7 h^{-1}\text{Mpc}$, $5 h^{-1}\text{Mpc}$, $8 h^{-1}\text{Mpc}$, $13 h^{-1}\text{Mpc}$, and $40 h^{-1}\text{Mpc}$ respectively. The color-dependent bias in V6 is independent of scale on all scales probed (above $16 h^{-1}\text{Mpc}$). In summary, our data shows the luminosity dependence of clustering to be constant down to cosmologically quite small scales, whereas the color dependence of clustering is slightly scale dependent.

In comparison to previous work [42, 50], it is perhaps surprising to see as little scale dependence as we do — [50] find the luminosity-dependent bias to vary with scale (see their Fig. 4), in contrast to what we find here. The measurement of luminosity-dependent bias in [42] agrees more closely with our observation of scale independence, but their Fig. 10 indicates that we might expect to see scale dependence of the luminosity-dependent bias in the most luminous samples. However, we measure the bias in our most luminous samples (in V6) at $16 - 134 h^{-1}\text{Mpc}$, well above the range probed in [42], so there is no direct conflict here. Additionally, Fig. 13 of [42] and Fig. 10 of [50] show that correlation functions of red and blue galaxies have significantly different slopes, implying that the color-dependent bias should be strongly scale-dependent on $0.1 - 10 h^{-1}\text{Mpc}$ scales. However, the points $> 1 h^{-1}\text{Mpc}$ in these plots (the range comparable to the scales we probe here) do not appear strongly scale dependent, so our results are not inconsistent with these correlation function measurements.

3. How bias depends on luminosity

Our next step is to calculate the relative bias parameter b/b_* (the bias relative to L_* galaxies) as a function of luminosity by combining the measured values of b_{rel} between the different pairs of luminosity bins. This function has been measured previously using SDSS power spectra [58] at length scales of $\sim 60 h^{-1}\text{Mpc}$ as well as SDSS [42, 50] and 2dFGRS [57] correlation functions at length scales of $\sim 1 h^{-1}\text{Mpc}$ — here we measure it at length scales of $\sim 20 h^{-1}\text{Mpc}$.

The bias of each luminosity bin relative to the central bin L4 is given by

$$\begin{aligned} \frac{b_1}{b_4} &= b_{12}b_{23}b_{34}, & \frac{b_2}{b_4} &= b_{23}b_{34}, & \frac{b_3}{b_4} &= b_{34}, \\ \frac{b_7}{b_4} &= \frac{1}{b_{45}b_{56}b_{67}}, & \frac{b_6}{b_4} &= \frac{1}{b_{45}b_{56}}, & \frac{b_5}{b_4} &= \frac{1}{b_{45}} \end{aligned} \quad (24)$$

where $b_{\alpha\beta}$ denotes the measured value of b_{rel} between luminosity bins L_α and L_β using all galaxies and b_α denotes the bias of galaxies in luminosity bin L_α relative to the dark matter. For each pairwise comparison, we choose the value of b_{rel} calculated at the resolution where the

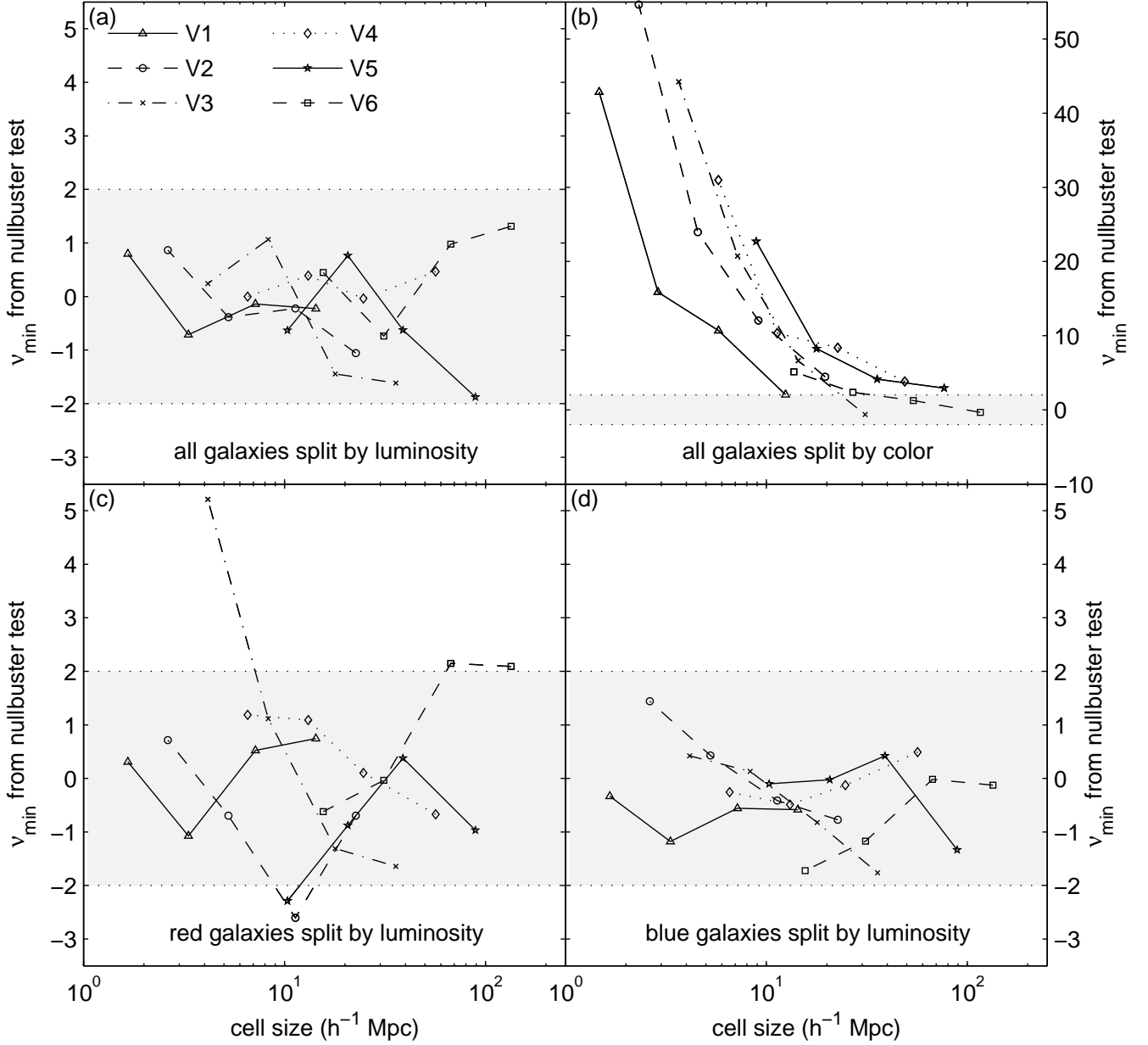


Figure 6: Null-buster results for pairwise comparisons. ν_{\min} measures the number of sigmas at which deterministic linear biasing can be ruled out as a model of relative bias between the two samples being compared. Shaded areas indicate $|\nu_{\min}| < 2$, where data is consistent with the null hypothesis at the 2σ level. Four different types of pairwise comparison are illustrated: (a) luminous vs. dim, (b) red vs. blue, (c) luminous red vs. dim red, and (d) luminous blue vs. dim blue. The different symbols denote the different comparison volumes V1-V6. The luminosity-dependent bias (a, c, d) is consistent with deterministic linear biasing but color-dependent bias (b) is not.

cell size is closest to $20 h^{-1} \text{Mpc}$, as illustrated in Fig. 5. (Since we see no evidence for scale dependence of b_{rel} for the luminosity-dependent bias, this choice does not strongly influence the results.)

To compute the error bars on b_{α}/b_4 , we rewrite equation (24) as a linear matrix equation using the logs of the bias values:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \log b_1/b_4 \\ \log b_2/b_4 \\ \log b_3/b_4 \\ \log b_5/b_4 \\ \log b_6/b_4 \\ \log b_7/b_4 \end{pmatrix} = \begin{pmatrix} \log b_{12} \\ \log b_{23} \\ \log b_{34} \\ \log b_{45} \\ \log b_{56} \\ \log b_{67} \end{pmatrix}, \quad (25)$$

or $\mathbf{A} \mathbf{b}_{\log} = \mathbf{b}_{\log, \text{rel}}$, where $\mathbf{b}_{\log, \text{rel}}$ is a vector of the log of

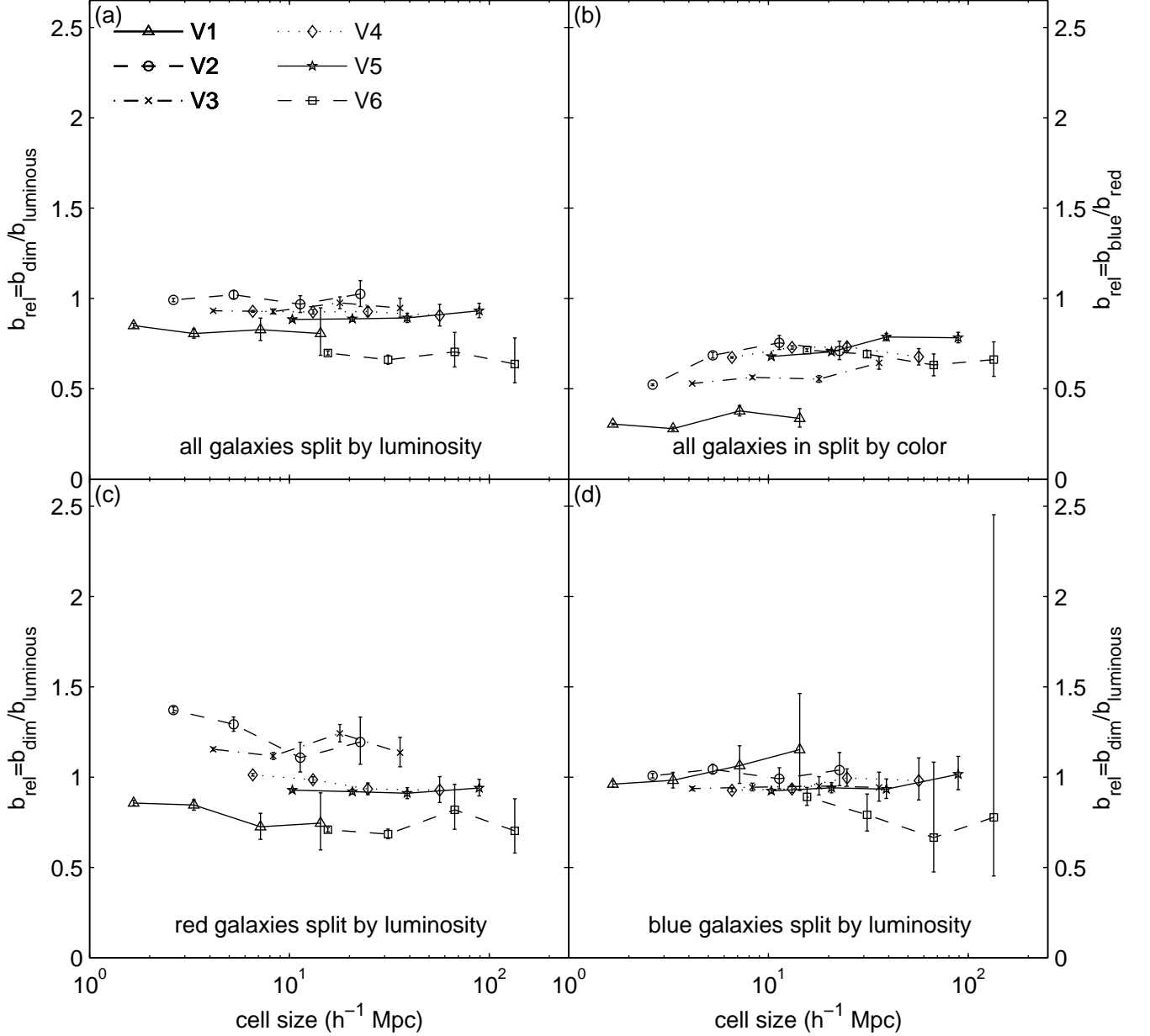


Figure 7: Relative bias b_{rel} between pairwise samples from null-buster analysis, revealing no significant scale dependence of luminosity-dependent bias on scales $\gg 5h^{-1}\text{Mpc}$, but a slight scale dependence of color-dependent bias. Four different types of pairwise comparison are illustrated: (a) luminous vs. dim, (b) red vs. blue, (c) luminous red vs. dim red, and (d) luminous blue vs. dim blue. The different symbols denote the different comparison volumes V1-V6.

our relative bias measurements $b_{\alpha\beta}$, \mathbf{b}_{\log} is a vector of the log of the bias values b_{α}/b_4 , and \mathbf{A} is the matrix relating them. Since each b_{rel} measurement is made in a separate volume, they are independent and the covariance matrix of $\mathbf{b}_{\log, \text{rel}}$ is given by

$$\Sigma_{\log, \text{rel}} = \text{diag} \left(\frac{\sigma_{12}^2}{b_{12}^2}, \frac{\sigma_{23}^2}{b_{23}^2}, \frac{\sigma_{34}^2}{b_{34}^2}, \frac{\sigma_{45}^2}{b_{45}^2}, \frac{\sigma_{56}^2}{b_{56}^2}, \frac{\sigma_{67}^2}{b_{67}^2} \right) \quad (26)$$

where $\sigma_{\alpha\beta}$ is the uncertainty on $b_{\alpha\beta}$, calculated as described in Section IIID. We invert equation (25) to give

\mathbf{b}_{\log} :

$$\mathbf{b}_{\log} = \mathbf{A}^{-1} \mathbf{b}_{\log, \text{rel}}, \quad (27)$$

with the covariance matrix for \mathbf{b}_{\log} given by

$$\Sigma_{\log} = \left(\mathbf{A}^T \Sigma_{\log, \text{rel}}^{-1} \mathbf{A} \right)^{-1}. \quad (28)$$

We then fit our data with the model used in [57]: $b(M)/b_* = a_1 + a_2(L/L_*)$, parameterized by $\mathbf{a} \equiv$

(a_1, a_2) . Here M is the central absolute magnitude of the bin, L is the corresponding luminosity, and $M_* = -20.83$. We use a weighted least-squares fit that is linear in the parameters (a_1, a_2) — that is, we solve the matrix equation

$$\begin{pmatrix} b_1/b_4 \\ b_2/b_4 \\ b_3/b_4 \\ b_4/b_4 \\ b_5/b_4 \\ b_6/b_4 \end{pmatrix} = \begin{pmatrix} 1 & L_1/L_4 \\ 1 & L_2/L_4 \\ 1 & L_3/L_4 \\ 1 & L_4/L_4 \\ 1 & L_5/L_4 \\ 1 & L_6/L_4 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad (29)$$

or $\mathbf{b} = \mathbf{X}\mathbf{a}$, where \mathbf{b} is a vector of the bias values b_α/b_4 and \mathbf{X} is the matrix representing our model. We solve for \mathbf{a} using

$$\mathbf{a} = (\mathbf{X}^T \mathbf{\Sigma} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{b}. \quad (30)$$

Here $\mathbf{\Sigma}$ is the covariance matrix of \mathbf{b} , given by

$$\mathbf{\Sigma} = \mathbf{B} \mathbf{\Sigma}_{\log} \mathbf{B}^T, \quad (31)$$

where $\mathbf{\Sigma}_{\log}$ is given by equation (28) and $\mathbf{B} \equiv \text{diag}(\mathbf{b})$. This procedure gives us the best-fit values for the parameters a_1 and a_2 , accounting for the correlations between the data points that are induced we compute the bias values \mathbf{b} from our relative bias measurements. We then normalize the model such that $b(M_*)/b_* = 1$.

Figure 8 shows a plot of b/b_* vs. M : results for all galaxies are plotted with black open circles, our best-fit model is shown by the solid line, the best-fit model from [57] is shown by the gray dashed line, and the best-fit model from [58] is shown by the dotted line. The error bars represent the diagonal elements of $\mathbf{\Sigma}$ from equation 31. Our model, with $(a_1, a_2) = (0.848, 0.152)$, agrees extremely well with the model from [57], with $(a_1, a_2) = (0.85, 0.15)$. This agreement is quite remarkable since we use data from a different survey and analyze it with a completely different technique.

We also use equation (27) to calculate b/b_* vs. M for red and blue galaxies separately. To normalize the red-galaxy data points to the all-galaxy data points in Fig. 8, we use best-fit values of σ_1^2 from the likelihood analysis described in Sec. III E: σ_1 from the comparison of dimmer and more luminous galaxies in V3 gives the normalization for all galaxies in L4, and σ_1 from the comparison of blue and red galaxies in L4 gives the normalization for red galaxies in L4, so the normalization factor for the red points is the square root of the ratio between the best-fit σ_1^2 values for these 2 comparisons. The blue points are then normalized relative to the red points using the measured value of b_{rel} from the null-buster comparison of blue and red galaxies in L4. Thus, the shapes of the red and blue curves are determined using the luminosity-dependent bias, which is independent of scale as seen in Fig. 7(c) and (d), but their relative normalization is determined by the color-dependent bias, which has some slight scale dependence as discussed in Section IV A 2.

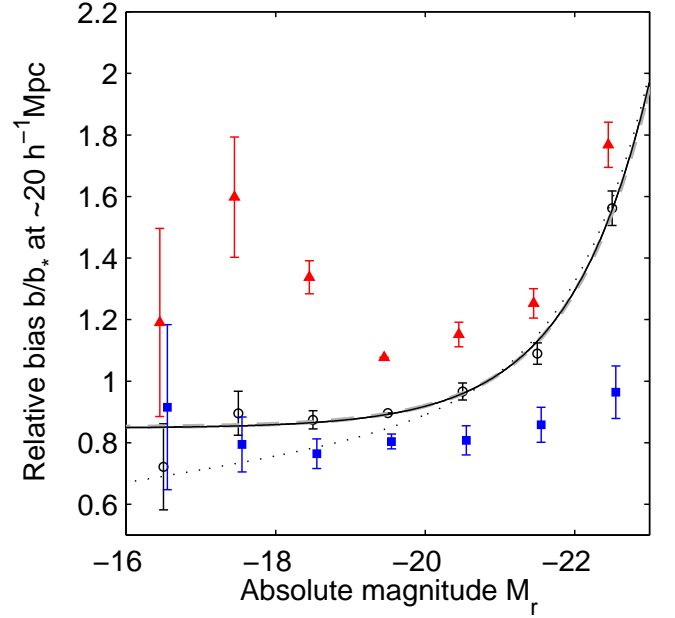


Figure 8: (Color online) Luminosity dependence of bias for all (open circles), red (solid triangles), and blue (solid squares) galaxies at a cell size of $\sim 20 h^{-1} \text{Mpc}$ from null-buster results. The solid line is a model fit to the all-galaxy data points, the gray dashed line shows the model from [57], and the dotted line shows the model from [58]. (Likelihood results look similar but with larger error bars.)

This means that making this plot at a different scale size would simply induce a small shift in the relative red/blue normalization but would not impact the shapes of the curves.

Splitting the luminosity dependence of the bias by color reveals some interesting features. The bias of the blue galaxies shows only a weak dependence on luminosity, and both luminous ($M \sim -22$) and dim ($M \sim -17$) red galaxies have slightly higher bias than moderately bright ($M \sim -20 \sim M_*$) red galaxies. The previously observed luminosity dependence of bias, with a weak dependence dimmer than L_* and a strong increase above L_* , is thus quite sensitive to the color selection: the lower luminosity bins contain mostly blue galaxies and thus show weak luminosity dependence, whereas the more luminous bins are dominated by red galaxies which drive the observed trend of more luminous galaxies being more strongly biased. It is instructive to compare these results with the mean local overdensity in color-magnitude space, as in Fig. 2 of [41]. Although our bias measurements are necessarily much coarser, it can be seen that the bias is strongest where the overdensity is largest, as might be expected.

Indications of these differing trends for red and blue galaxies have been observed in previous work: an early hint of the upturn in the bias for dim red galaxies was seen in [35], and is consistent with the halo model-based predictions in [13], although not with the predictions in

[16]. There is some inconsistency between our results and [42, 50] regarding the dim red galaxies: they find that dim red galaxies exhibit the strongest clustering on scales $< 1 h^{-1}\text{Mpc}$ and luminous red galaxies exhibit the strongest clustering on larger scales, as shown in Fig. 14 of [42] and Fig. 11 of [50]. However, we find the dim red galaxies to have higher bias than L_* red galaxies at all the scales we probe ($2 - 40 h^{-1}\text{Mpc}$ in this case).

Recent measurements of higher order clustering statistics [52] find the same trends in the clustering strengths of red and blue galaxies, although they indicate that their linear bias measurement (which should be comparable to ours) shows the opposite trends — little luminosity dependence for red galaxies and a slight monotonic increase for blue galaxies. This may be at least in part because their luminosity range is much narrower than ours. Some of the previous studies have also reported a somewhat stronger luminosity dependence of blue galaxy clustering than we have measured here.

B. Likelihood Results

To study the luminosity dependence, color dependence and stochasticity of bias in more detail, we also apply the maximum likelihood method described in Sec. III E to all of the same pairs of samples used in the null-buster test. Due to constraints on computing power and memory, we perform these calculations for only three values of the cell size L rather than four, dropping the highest resolution (smallest cell size) shown in Fig. 4. The likelihood analysis makes a few additional assumptions, but provides a valuable cross-check and also a measurement of the parameter r_{rel} which encodes the stochasticity and nonlinearity of the relative bias.

For each pair of samples, the likelihood function given in equation (22) is maximized with respect to the parameters σ_1^2 , b_{rel} , and r_{rel} and marginalized over σ_1^2 to determine the best-fit values of b_{rel} and r_{rel} , with uncertainties defined by the $\Delta(2\ln\mathcal{L}) = 1$ contour in the $b_{\text{rel}}-r_{\text{rel}}$ plane. Additionally, 100 mock data sets are generated using the best-fit parameter values from the data as described in Sec. III E, and the likelihood minimization is then applied to the mock samples as well. The uncertainties on b_{rel} and r_{rel} calculated from the likelihood contours agree well with the standard deviation of the best-fit values from the mock data sets — below we use the uncertainties from the likelihood contours. The values of b_{rel} are consistent with those determined using the null-buster test shown in Fig. 7.

Figure 9 shows the best-fit values of r_{rel} as a function of cell size L . For the comparisons between neighboring luminosity bins, the results are consistent with $r_{\text{rel}} = 1$. On the other hand, the comparisons between red and blue galaxies give $r_{\text{rel}} < 1$, with smaller cell sizes L giving smaller values of r_{rel} . This confirms the null-buster result that the luminosity-dependent bias can be accurately modeled using simple deterministic linear bias

but color-dependent bias demands a more complicated model. Also, r_{rel} for the color-dependent bias is seen to depend on scale but not on luminosity. In contrast, b_{rel} (both in the null-buster and likelihood analyses) depends on luminosity but is at most weakly dependent on scale.

To summarize, we find that the simple, deterministic model is a good fit for the luminosity-dependent bias, but the color-dependent bias shows evidence for stochasticity and/or nonlinearity which increases in strength towards smaller scales. These results are consistent with previous detections of stochasticity/nonlinearity in spectral-type-dependent bias [23, 24, 28], and also agree with [43] which measures significant stochasticity between galaxies of different color or spectral type, but not between galaxies of different luminosities. Our plots in Fig. 7(b) and Fig. 9(b) showing the scale dependence of b_{rel} and r_{rel} for the color-dependent bias agree quite well with the plots of b_{var} and r_{lin} in [43] (their Fig. 11), implying that these results are quite robust since our analysis uses a different data set, employs different methods, and makes different assumptions. However, our results are somewhat at odds with the cross-correlation measurement in [42], which indicates that the cross-correlation between red and blue galaxies shown in their Fig. 24 closely follows the geometric means of their auto-correlations, implying that the cross-correlation coefficient r_{rel} is close to 1.

V. CONCLUSIONS

To shed further light on how galaxies trace matter, we have quantified how different types of galaxies trace each other. We have analyzed the relative bias between pairs of volume-limited galaxy samples of different luminosities and colors using counts-in-cells at varying length scales. This method is most sensitive to length scales between those probed by correlation function and power spectrum methods, and makes point-by-point comparisons of the density fields rather than using ratios of moments, thereby eliminating sample variance and obtaining a local rather than global measure of the bias. We applied a null-buster test on each pair of subsamples to determine if the relative bias was consistent with deterministic linear biasing, and we also performed a maximum-likelihood analysis to find the best-fit parameters for a simple stochastic biasing model.

A. Biasing results

Our primary results are:

- The luminosity-dependent bias for red galaxies is significantly different from that of blue galaxies: the bias of blue galaxies shows only a weak dependence on luminosity, whereas both luminous and dim red galaxies have higher bias than moderately bright (L_*) red galaxies.

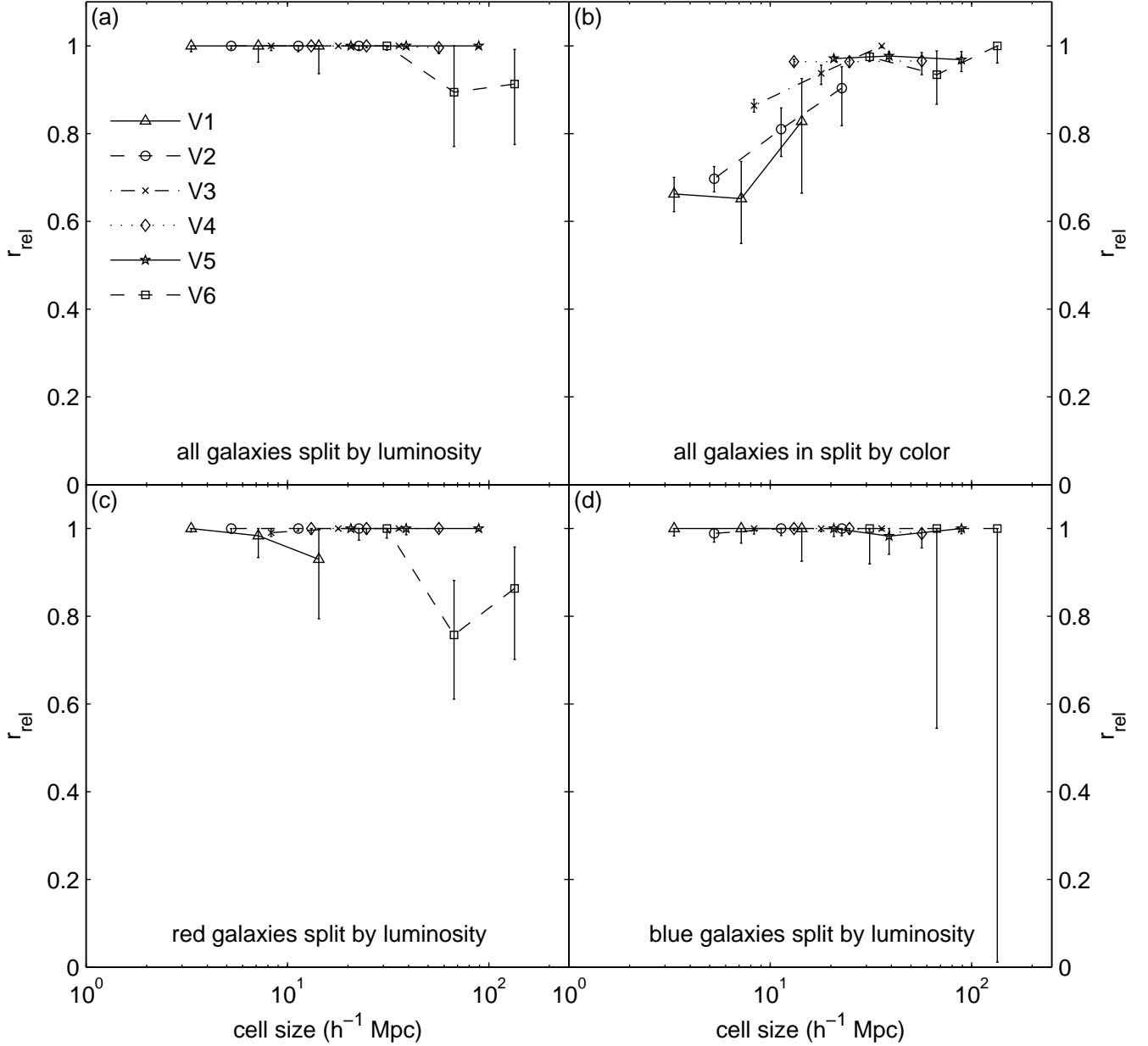


Figure 9: The best-fit values of the relative cross-correlation coefficient r_{rel} between pairwise samples. Four different types of pairwise comparison are illustrated: (a) luminous vs. dim, (b) red vs. blue, (c) luminous red vs. dim red, and (d) luminous blue vs. dim blue. The different symbols denote the different comparison volumes V1-V6.

- Both of our analysis methods indicate that the simple, deterministic model is a good fit for the luminosity-dependent bias, but that the color-dependent bias is more complicated, showing strong evidence for stochasticity and/or nonlinearity on scales $\ll 10 h^{-1} \text{ Mpc}$.
- The luminosity-dependent bias is consistent with being scale-independent over the range of scales probed here ($2 - 160 h^{-1} \text{ Mpc}$). The color-dependent bias depends on luminosity but is only weakly scale-dependent, while the cross-correlation

coefficient r_{rel} depends on scale but not luminosity, giving smaller r_{rel} values at smaller scales.

These results are encouraging from the perspective of using galaxy clustering to measure cosmological parameters: simple scale-independent linear biasing appears to be a good approximation on the $\gtrsim 60 h^{-1} \text{ Mpc}$ scales used in many recent cosmological studies (e.g., [79, 154, 155]). However, further quantification of small residual effects will be needed to do full justice to the precision of next-generation data sets on the horizon. Moreover, our results regarding color sensitivity suggest that more de-

tailed bias studies are worthwhile for luminous red galaxies, which have emerged a powerful cosmological probe because of their visibility at large distances and near-optimal number density [79, 132, 137], since color cuts are involved in their selection.

B. Implications for galaxy formation

What can these results tell us about galaxy formation in the context of the halo model? First of all, as discussed in [42], the large bias of the faint red galaxies can be explained by the fact that such galaxies tend to be satellites in high mass halos, which are more strongly clustered than low mass halos. Previous studies have found that central galaxies in low-mass halos are preferentially blue, central galaxies in high mass halos tend to be red, and that the luminosity of the central galaxy is strongly correlated with the halo mass [45, 98]. Our observed lack of luminosity dependence of the bias for blue galaxies would then be a reflection of the correlation between luminosity and halo mass being weaker for blue galaxies than for red ones. Additional work is needed to study this quantitatively and compare it with theoretical predictions from galaxy formation models.

The detection of stochasticity between red and blue galaxies may imply that red and blue galaxies tend to live in different halos — a study of galaxy groups in SDSS [49] recently presented evidence supporting this, but this is at odds with the cross-correlation measurement in [42], which implies that blue and red galaxies are well-mixed within halos. The fact that the stochasticity is strongest at small scales suggests that this effect is due to the 1-halo term, i.e. arising from pairs of galaxies in the same halo, although some amount of stochasticity persists even for large scales. However, the halo model implications for stochasticity have not been well-studied to date.

In summary, our results on galaxy biasing and future work along these lines should be able to deepen our understanding of both cosmology (by quantifying system-

atic uncertainties) and galaxy formation.

Acknowledgments

The authors wish to thank Andrew Hamilton for work with the Mangle software, and Daniel Eisenstein, David Hogg, Taka Matsubara, Ryan Scranton, Ramin Skibba, and Simon White for helpful comments. This work was supported by NASA grants NAG5-11099 and NNG06GC55G, NSF grants AST-0134999 and 0607597, the Kavli Foundation, and fellowships from the David and Lucile Packard Foundation and the Research Corporation.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, Cambridge University, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

-
- [1] E. Hubble and M. L. Humason, *ApJ* **74**, 43 (1931).
 - [2] M. Davis and M. J. Geller, *ApJ* **208**, 13 (1976).
 - [3] A. Dressler, *ApJ* **236**, 351 (1980).
 - [4] M. Postman and M. J. Geller, *ApJ* **281**, 95 (1984).
 - [5] A. J. S. Hamilton, *ApJ* **331**, L59 (1988).
 - [6] S. D. M. White, R. B. Tully, and M. Davis, *ApJ* **333**, L45 (1988).
 - [7] G. Boerner, Z.-G. Deng, X.-Y. Xia, and Y.-Y. Zhou, *Ap&SS* **180**, 47 (1991).
 - [8] C. Park, M. S. Vogeley, M. J. Geller, and J. P. Huchra, *ApJ* **431**, 569 (1994).
 - [9] J. Loveday, S. J. Maddox, G. Efstathiou, and B. A. Peterson, *ApJ* **442**, 457 (1995), [astro-ph/9410018](http://arxiv.org/abs/astro-ph/9410018).
 - [10] S. Hermit, B. X. Santiago, O. Lahav, M. A. Strauss, M. Davis, A. Dressler, and J. P. Huchra, *MNRAS* **283**, 709 (1996), [astro-ph/9608001](http://arxiv.org/abs/astro-ph/9608001).
 - [11] U. Seljak, *MNRAS* **325**, 1359 (2001), [astro-ph/0009016](http://arxiv.org/abs/astro-ph/0009016).
 - [12] A. Tasitsiomi, A. V. Kravtsov, R. H. Wechsler, and J. R. Primack, *ApJ* **614**, 533 (2004), [astro-ph/0404168](http://arxiv.org/abs/astro-ph/0404168).
 - [13] F. C. van den Bosch, X. Yang, and H. J. Mo, *MNRAS* **340**, 771 (2003), [astro-ph/0210495](http://arxiv.org/abs/astro-ph/0210495).
 - [14] D. H. Weinberg, R. Davé, N. Katz, and L. Hernquist, *ApJ* **601**, 1 (2004), [astro-ph/0212356](http://arxiv.org/abs/astro-ph/0212356).
 - [15] A. A. Berlind, M. R. Blanton, D. W. Hogg, D. H. Weinberg, R. Davé, D. J. Eisenstein, and N. Katz, *ApJ* **629**, 625 (2005), [astro-ph/0406633](http://arxiv.org/abs/astro-ph/0406633).
 - [16] A. Cooray, *MNRAS* **363**, 337 (2005), [astro-ph/0505421](http://arxiv.org/abs/astro-ph/0505421).
 - [17] C. Conroy, R. H. Wechsler, and A. V. Kravtsov, *ApJ* **647**, 201 (2006), [astro-ph/0512234](http://arxiv.org/abs/astro-ph/0512234).
 - [18] J. L. Tinker, P. Norberg, D. H. Weinberg, and M. S. Warren, [astro-ph/0603543](http://arxiv.org/abs/astro-ph/0603543) (2006).
 - [19] A. Slosar, U. Seljak, and A. Tasitsiomi, *MNRAS* **366**,

- 1455 (2006), astro-ph/0507203.
- [20] R. K. Sheth, A. J. Connolly, and R. Skibba, astro-ph/0511773 (2005).
- [21] L. Guzzo, M. A. Strauss, K. B. Fisher, R. Giovanelli, and M. P. Haynes, *ApJ* **489**, 37 (1997), astro-ph/9706150.
- [22] A. Dressler, A. J. Oemler, W. J. Couch, I. Smail, R. S. Ellis, A. Barger, H. Butcher, B. M. Poggianti, and R. M. Sharples, *ApJ* **490**, 577 (1997), astro-ph/9707232.
- [23] M. Tegmark and B. C. Bromley, *ApJ* **518**, L69 (1999), astro-ph/9809324.
- [24] M. Blanton, *ApJ* **544**, 63 (2000), astro-ph/0003228.
- [25] M. J. I. Brown, R. L. Webster, and B. J. Boyle, *MNRAS* **317**, 782 (2000), astro-ph/0005329.
- [26] D. S. Madgwick, E. Hawkins, O. Lahav, S. Maddox, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, et al., *MNRAS* **344**, 847 (2003), astro-ph/0303668.
- [27] T. Goto, C. Yamauchi, Y. Fujita, S. Okamura, M. Sekiguchi, I. Smail, M. Bernardi, and P. L. Gomez, *MNRAS* **346**, 601 (2003), astro-ph/0312043.
- [28] E. Conway, S. Maddox, V. Wild, J. A. Peacock, E. Hawkins, P. Norberg, D. S. Madgwick, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, et al., *MNRAS* **356**, 456 (2005), astro-ph/0404276.
- [29] B. Meneux, O. Le Fèvre, L. Guzzo, A. Pollo, A. Cappi, O. Ilbert, A. Iovino, C. Marinoni, H. J. McCracken, D. Bottini, et al., *A&A* **452**, 387 (2006), astro-ph/0511656.
- [30] A. J. Ross, R. J. Brunner, and A. D. Myers, *ApJ* **649**, 48 (2006), astro-ph/0605748.
- [31] S. Phleps, C. Wolf, J. A. Peacock, K. Meisenheimer, and E. van Kampen, astro-ph/0612066 (2006).
- [32] C. N. A. Willmer, L. N. da Costa, and P. S. Pellegrini, *AJ* **115**, 869 (1998), astro-ph/9803118.
- [33] R. G. Carlberg, H. K. C. Yee, S. L. Morris, H. Lin, P. B. Hall, D. R. Patton, M. Sawicki, and C. W. Shepherd, *ApJ* **563**, 736 (2001), astro-ph/0106506.
- [34] I. Zehavi, M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo, M. A. Strauss, S. F. Anderson, J. Annis, N. A. Bahcall, M. Bernardi, et al., *ApJ* **571**, 172 (2002), astro-ph/0106476.
- [35] P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, D. Madgwick, O. Lahav, S. Cole, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, et al., *MNRAS* **332**, 827 (2002), astro-ph/0112043.
- [36] D. W. Hogg, M. R. Blanton, D. J. Eisenstein, J. E. Gunn, D. J. Schlegel, I. Zehavi, N. A. Bahcall, J. Brinkmann, I. Csabai, D. P. Schneider, et al., *ApJ* **585**, L5 (2003), astro-ph/0212085.
- [37] T. Budavári, A. J. Connolly, A. S. Szalay, I. Szapudi, I. Csabai, R. Scranton, N. A. Bahcall, J. Brinkmann, D. J. Eisenstein, J. A. Frieman, et al., *ApJ* **595**, 59 (2003), astro-ph/0305603.
- [38] D. W. Hogg, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. J. Schlegel, J. E. Gunn, T. A. McKay, H.-W. Rix, N. A. Bahcall, J. Brinkmann, et al., *ApJ* **601**, L29 (2004), astro-ph/0307336.
- [39] A. L. Coil, M. Davis, D. S. Madgwick, J. A. Newman, C. J. Conselice, M. Cooper, R. S. Ellis, S. M. Faber, D. P. Finkbeiner, P. Guhathakurta, et al., *ApJ* **609**, 525 (2004), astro-ph/0305586.
- [40] C. Marinoni, O. Le Fèvre, B. Meneux, A. Iovino, A. Pollo, O. Ilbert, G. Zamorani, L. Guzzo, A. Mazure, R. Scaramella, et al., *A&A* **442**, 801 (2005), astro-ph/0506561.
- [41] M. R. Blanton, D. Eisenstein, D. W. Hogg, D. J. Schlegel, and J. Brinkmann, *ApJ* **629**, 143 (2005), astro-ph/0310453.
- [42] I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, et al., *ApJ* **630**, 1 (2005), astro-ph/0408569.
- [43] V. Wild, J. A. Peacock, O. Lahav, E. Conway, S. Maddox, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, et al., *MNRAS* **356**, 247 (2005), astro-ph/0404275.
- [44] D. J. Croton, G. R. Farrar, P. Norberg, M. Colless, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, et al., *MNRAS* **356**, 1155 (2005), astro-ph/0407537.
- [45] X. Yang, H. J. Mo, Y. P. Jing, and F. C. van den Bosch, *MNRAS* **358**, 217 (2005), astro-ph/0410114.
- [46] X. Yang, H. J. Mo, F. C. van den Bosch, S. M. Weinmann, C. Li, and Y. P. Jing, *MNRAS* **362**, 711 (2005), astro-ph/0504477.
- [47] E. Gaztañaga, P. Norberg, C. M. Baugh, and D. J. Croton, *MNRAS* **364**, 620 (2005), astro-ph/0506249.
- [48] M. R. Blanton, D. Eisenstein, D. W. Hogg, and I. Zehavi, *ApJ* **645**, 977 (2006), astro-ph/0411037.
- [49] S. M. Weinmann, F. C. van den Bosch, X. Yang, and H. J. Mo, *MNRAS* **366**, 2 (2006), astro-ph/0509147.
- [50] C. Li, G. Kauffmann, Y. P. Jing, S. D. M. White, G. Börner, and F. Z. Cheng, *MNRAS* **368**, 21 (2006), astro-ph/0509873.
- [51] R. Skibba, R. K. Sheth, A. J. Connolly, and R. Scranton, *MNRAS* **369**, 68 (2006), astro-ph/0512463.
- [52] D. J. Croton, P. Norberg, E. Gaztañaga, and C. M. Baugh, astro-ph/0611313 (2006).
- [53] C. Li, Y. P. Jing, G. Kauffmann, G. Boerner, X. Kang, and L. Wang, astro-ph/0701218 (2007).
- [54] L. Guzzo, O. Le Fèvre, B. Meneux, A. Pollo, C. Marinoni, A. Cappi, O. Cucciati, B. Garilli, A. Iovino, H. J. McCracken, et al., astro-ph/0701273 (2007).
- [55] T. Ichikawa, R. Suzuki, C. Tokoku, Y. Katsuno-Uchimoto, M. Konishi, T. Yoshikawa, M. Kajisawa, M. Ouchi, T. Hamana, M. Akiyama, et al., astro-ph/0701820 (2007).
- [56] C. Benoist, A. Cappi, L. N. da Costa, S. Maurogordato, F. R. Bouchet, and R. Schaeffer, *ApJ* **514**, 563 (1999).
- [57] P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, J. A. Peacock, S. Cole, C. S. Frenk, J. Bland-Hawthorn, T. Bridges, R. Cannon, et al., *MNRAS* **328**, 64 (2001), astro-ph/0105500.
- [58] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, et al., *ApJ* **606**, 702 (2004), astro-ph/0310725.
- [59] D. J. Eisenstein, M. Blanton, I. Zehavi, N. Bahcall, J. Brinkmann, J. Loveday, A. Meiksin, and D. Schneider, *ApJ* **619**, 178 (2005), astro-ph/0411559.
- [60] I. Zehavi, D. J. Eisenstein, R. C. Nichol, M. R. Blanton, D. W. Hogg, J. Brinkmann, J. Loveday, A. Meiksin, D. P. Schneider, and M. Tegmark, *ApJ* **621**, 22 (2005), astro-ph/0411557.
- [61] M. Ouchi, T. Hamana, K. Shimasaku, T. Yamada, M. Akiyama, N. Kashikawa, M. Yoshida, K. Aoki, M. Iye, T. Saito, et al., *ApJ* **635**, L117 (2005), astro-

- ph/0508083.
- [62] P. D. Allen, L. A. Moustakas, G. Dalton, E. MacDon-ald, C. Blake, L. Clewley, C. Heymans, and G. Wegner, *MNRAS* **360**, 1244 (2005), astro-ph/0505015.
 - [63] U. Seljak, A. Makarov, R. Mandelbaum, C. M. Hirata, N. Padmanabhan, P. McDonald, M. R. Blanton, M. Tegmark, N. A. Bahcall, and J. Brinkmann, *Phys. Rev. D* **71**, 043511 (2005), astro-ph/0406594.
 - [64] A. Pollo, L. Guzzo, O. Le Fèvre, B. Meneux, A. Cappi, P. Franzetti, A. Iovino, H. J. McCracken, C. Marioni, G. Zamorani, et al., *A&A* **451**, 409 (2006), astro-ph/0512429.
 - [65] A. Cooray and M. Ouchi, *MNRAS* **369**, 1869 (2006), astro-ph/0604479.
 - [66] B. Pandey and S. Bharadwaj, astro-ph/0701615 (2007).
 - [67] M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, et al., *MNRAS* **328**, 1039 (2001), astro-ph/0106498.
 - [68] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, et al., *AJ* **120**, 1579 (2000), astro-ph/0006396.
 - [69] C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles, F. J. Castander, A. J. Connolly, D. J. Eisenstein, J. A. Frieman, G. S. Hennessy, et al., *AJ* **123**, 485 (2002).
 - [70] P. Schechter, *ApJ* **203**, 297 (1976).
 - [71] U.-L. Pen, *ApJ* **504**, 601 (1998), astro-ph/9711180.
 - [72] M. Tegmark and P. J. E. Peebles, *ApJ* **500**, L79+ (1998), astro-ph/9804067.
 - [73] A. Dekel and O. Lahav, *ApJ* **520**, 24 (1999), astro-ph/9806193.
 - [74] T. Matsubara, *ApJ* **525**, 543 (1999), astro-ph/9906029.
 - [75] P. Simon, M. Hetterscheidt, M. Schirmer, T. Erben, P. Schneider, C. Wolf, and K. Meisenheimer, *A&A* **461**, 861 (2007), astro-ph/0606622.
 - [76] P. Simon, *A&A* **430**, 827 (2005), astro-ph/0409435.
 - [77] R. Casas-Miranda, H. J. Mo, R. K. Sheth, and G. Boerner, *MNRAS* **333**, 730 (2002), astro-ph/0105008.
 - [78] U. Seljak and M. S. Warren, *MNRAS* **355**, 129 (2004), astro-ph/0403698.
 - [79] M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp, et al., *Phys. Rev. D* **74**, 123507 (2006), astro-ph/0608632.
 - [80] C. Blake, A. Collister, S. Bridle, and O. Lahav, *MNRAS* **374**, 1527 (2007), astro-ph/0605303.
 - [81] N. Padmanabhan, D. J. Schlegel, U. Seljak, A. Makarov, N. A. Bahcall, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. P. Finkbeiner, J. E. Gunn, et al., astro-ph/0605302 (2006).
 - [82] W. J. Percival, L. Verde, and J. A. Peacock, *MNRAS* **347**, 645 (2004), astro-ph/0306511.
 - [83] K. Abazajian, Z. Zheng, I. Zehavi, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, N. A. Bahcall, J. Brinkmann, D. P. Schneider, et al., *ApJ* **625**, 613 (2005), astro-ph/0408003.
 - [84] Z. Zheng and D. H. Weinberg, astro-ph/0512071 (2005).
 - [85] O. Moeller, M. Kitzbichler, and P. Natarajan, astro-ph/0607032 (2006).
 - [86] W. J. Percival, R. C. Nichol, D. J. Eisenstein, J. A. Frieman, M. Fukugita, J. Loveday, A. C. Pope, D. P. Schneider, A. S. Szalay, M. Tegmark, et al., astro-ph/0608636 (2006).
 - [87] J. R. Kristiansen, O. Elgaroy, and H. Dahle, astro-ph/0611761 (2006).
 - [88] U. Seljak, *MNRAS* **318**, 203 (2000), astro-ph/0001493.
 - [89] A. Cooray and R. Sheth, *Phys. Rep.* **372**, 1 (2002), astro-ph/0206508.
 - [90] P. Catelan, F. Lucchin, S. Matarrese, and C. Porciani, *MNRAS* **297**, 692 (1998), astro-ph/9708067.
 - [91] P. McDonald, *Phys. Rev. D* **74**, 103512 (2006), astro-ph/0609413.
 - [92] R. E. Smith, R. Scoccimarro, and R. K. Sheth, astro-ph/0609547 (2006).
 - [93] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, *MNRAS* **341**, 1311 (2003), astro-ph/0207664.
 - [94] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, and J. R. Primack, *ApJ* **609**, 35 (2004), astro-ph/0308519.
 - [95] A. V. Kravtsov, astro-ph/0607463 (2006).
 - [96] R. S. Somerville, G. Lemson, Y. Sigad, A. Dekel, G. Kauffmann, and S. D. M. White, *MNRAS* **320**, 289 (2001), astro-ph/9912073.
 - [97] A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, A. Jenkins, N. Katz, and C. G. Lacey, *ApJ* **593**, 1 (2003), astro-ph/0212357.
 - [98] Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, N. Katz, and C. G. Lacey, *ApJ* **633**, 791 (2005), astro-ph/0408564.
 - [99] L. Wang, C. Li, G. Kauffmann, and G. de Lucia, *MNRAS* **371**, 537 (2006), astro-ph/0603546.
 - [100] R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey, *MNRAS* **370**, 645 (2006), astro-ph/0511338.
 - [101] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida, *MNRAS* **365**, 11 (2006), astro-ph/0508046.
 - [102] C. M. Baugh, *Reports of Progress in Physics* **69**, 3101 (2006), astro-ph/0610031.
 - [103] J. A. Peacock and R. E. Smith, *MNRAS* **318**, 1144 (2000), astro-ph/0005010.
 - [104] A. A. Berlind and D. H. Weinberg, *ApJ* **575**, 587 (2002), astro-ph/0109001.
 - [105] E. Sefusatti and R. Scoccimarro, *Phys. Rev. D* **71**, 063001 (2005), astro-ph/0412626.
 - [106] X. Yang, H. J. Mo, and F. C. van den Bosch, *MNRAS* **339**, 1057 (2003), astro-ph/0207019.
 - [107] X. Yang, H. J. Mo, and F. C. van den Bosch, *ApJ* **638**, L55 (2006), astro-ph/0509626.
 - [108] G. Zhu, Z. Zheng, W. P. Lin, Y. P. Jing, X. Kang, and L. Gao, *ApJ* **639**, L5 (2006), astro-ph/0601120.
 - [109] A. Cooray, astro-ph/0601090 (2006).
 - [110] D. J. Croton, L. Gao, and S. D. M. White, *MNRAS* **374**, 1303 (2007), astro-ph/0605636.
 - [111] L. Gao and S. D. M. White, astro-ph/0611921 (2006).
 - [112] A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey, *MNRAS* **311**, 793 (2000), astro-ph/9903343.
 - [113] R. Scranton, *MNRAS* **339**, 410 (2003), astro-ph/0205517.

- [114] Y.-T. Lin, J. J. Mohr, and S. A. Stanford, *ApJ* **610**, 745 (2004), astro-ph/0402308.
- [115] A. A. Collister and O. Lahav, *MNRAS* **361**, 415 (2005), astro-ph/0412516.
- [116] J. L. Tinker, D. H. Weinberg, and M. S. Warren, *ApJ* **647**, 737 (2006), astro-ph/0603146.
- [117] U. Abbas and R. K. Sheth, *MNRAS* **372**, 1749 (2006), astro-ph/0601407.
- [118] M. R. Blanton, A. A. Berlind, and D. W. Hogg, astro-ph/0608353 (2006).
- [119] U. Abbas and R. K. Sheth, *MNRAS* **364**, 1327 (2005), astro-ph/0509593.
- [120] Y. Wang, X. Yang, H. J. Mo, F. C. van den Bosch, and Y. Chu, *MNRAS* **353**, 287 (2004), astro-ph/0404143.
- [121] M. Magliocchetti and C. Porciani, *MNRAS* **346**, 186 (2003), astro-ph/0304003.
- [122] J. E. Gunn, M. Carr, C. Rockosi, M. Sekiguchi, K. Berry, B. Elms, E. de Haas, Ž. Ivezić, G. Knapp, R. Lupton, et al., *AJ* **116**, 3040 (1998), astro-ph/9809085.
- [123] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger, L. N. Carey, G. R. Knapp, D. G. York, W. N. Boroski, et al., *AJ* **131**, 2332 (2006), astro-ph/0602326.
- [124] M. Fukugita, T. Ichikawa, J. E. Gunn, M. Doi, K. Shimazaki, and D. P. Schneider, *AJ* **111**, 1748 (1996).
- [125] J. R. Pier, J. A. Munn, R. B. Hindsley, G. S. Hennessy, S. M. Kent, R. H. Lupton, and Ž. Ivezić, *AJ* **125**, 1559 (2003), astro-ph/0211375.
- [126] R. Lupton, J. E. Gunn, Z. Ivezić, G. R. Knapp, and S. Kent, in *ASP Conf. Ser. 238: Astronomical Data Analysis Software and Systems X*, edited by F. R. Harnenden, Jr., F. A. Primini, and H. E. Payne (2001), pp. 269–+.
- [127] D. W. Hogg, D. P. Finkbeiner, D. J. Schlegel, and J. E. Gunn, *AJ* **122**, 2129 (2001), astro-ph/0106511.
- [128] J. A. Smith, D. L. Tucker, S. Kent, M. W. Richmond, M. Fukugita, T. Ichikawa, S.-i. Ichikawa, A. M. Jorgensen, A. Uomoto, J. E. Gunn, et al., *AJ* **123**, 2121 (2002), astro-ph/0201143.
- [129] Z. Ivezić, R. H. Lupton, D. Schlegel, B. Boroski, J. Adelman-McCarthy, B. Yanny, S. Kent, C. Stoughton, D. Finkbeiner, N. Padmanabhan, et al., *Astronomische Nachrichten* **325**, 583 (2004), astro-ph/0410195.
- [130] D. L. Tucker, S. Kent, M. W. Richmond, J. Annis, J. A. Smith, S. S. Allam, C. T. Rodgers, J. L. Stute, J. K. Adelman-McCarthy, J. Brinkmann, et al., *Astronomische Nachrichten* **327**, 821 (2006), astro-ph/0608575.
- [131] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, *ApJ* **500**, 525 (1998), astro-ph/9710327.
- [132] D. J. Eisenstein, J. Annis, J. E. Gunn, A. S. Szalay, A. J. Connolly, R. C. Nichol, N. A. Bahcall, M. Bernardi, S. Burles, F. J. Castander, et al., *AJ* **122**, 2267 (2001), astro-ph/0108153.
- [133] M. R. Blanton, H. Lin, R. H. Lupton, F. M. Maley, N. Young, I. Zehavi, and J. Loveday, *AJ* **125**, 2276 (2003), astro-ph/0105535.
- [134] A. Uomoto, S. A. Smee, and R. H. Barkhouser, in *Ground-based Instrumentation for Astronomy. Edited by Alan F. M. Moorwood and Iye Masanori. Proceedings of the SPIE, Volume 5492, pp. 1411-1422 (2004).*, edited by A. F. M. Moorwood and M. Iye (2004), pp. 1411–1422.
- [135] J. Adelman-McCarthy et al. (2007), submitted to *ApJS*.
- [136] M. R. Blanton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, D. Finkbeiner, M. Fukugita, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, et al., *AJ* **129**, 2562 (2005), astro-ph/0410166.
- [137] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, et al., *ApJ* **633**, 560 (2005), astro-ph/0501171.
- [138] A. J. S. Hamilton and M. Tegmark, *MNRAS* **349**, 115 (2004), astro-ph/0306324.
- [139] M. R. Blanton, J. Brinkmann, I. Csabai, M. Doi, D. Eisenstein, M. Fukugita, J. E. Gunn, D. W. Hogg, and D. J. Schlegel, *AJ* **125**, 2348 (2003), astro-ph/0205243.
- [140] M. L. Balogh, I. K. Baldry, R. Nichol, C. Miller, R. Bower, and K. Glazebrook, *ApJ* **615**, L101 (2004), astro-ph/0406266.
- [141] A. Mateus, L. Sodr , R. Cid Fernandes, G. Stasi ska, W. Schoenell, and J. M. Gomes, *MNRAS* **370**, 721 (2006), astro-ph/0511578.
- [142] I. K. Baldry, M. L. Balogh, R. G. Bower, K. Glazebrook, R. C. Nichol, S. P. Bamford, and T. Budavari, *MNRAS* **373**, 469 (2006), astro-ph/0607648.
- [143] R. Scranton et al., <http://lahmu.phyast.pit.edu/scranton/SDSSPix>.
- [144] M. E. C. Swanson, M. Tegmark, A. J. S. Hamilton, and J. C. Hill (2007), in preparation.
- [145] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, et al., *MNRAS* **362**, 505 (2005), astro-ph/0501174.
- [146] J. A. Peacock, *Cosmological Physics* (Cosmological Physics, by John A. Peacock, pp. 704. ISBN 052141072X. Cambridge, UK: Cambridge University Press, January 1999., 1999).
- [147] B. Novosyadlyj, R. Durrer, and V. N. Lukash, *A&A* **347**, 799 (1999), astro-ph/9811262.
- [148] M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, et al., *Phys. Rev. D* **69**, 103501 (2004), astro-ph/0310723.
- [149] M. Tegmark, *ApJ* **519**, 513 (1999), astro-ph/9809001.
- [150] T. J. Broadhurst, A. N. Taylor, and J. A. Peacock, *ApJ* **438**, 49 (1995), astro-ph/9406052.
- [151] R. K. Sheth and A. Diaferio, *MNRAS* **322**, 901 (2001), astro-ph/0009166.
- [152] S. Dodelson, L. Hui, and A. Jaffe, astro-ph/9712074 (1997).
- [153] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C. The art of scientific computing* (Cambridge: University Press, |c1992, 2nd ed., 1992).
- [154] A. G. S nchez, C. M. Baugh, W. J. Percival, J. A. Peacock, N. D. Padilla, S. Cole, C. S. Frenk, and P. Norberg, *MNRAS* **366**, 189 (2006), astro-ph/0507583.
- [155] D. N. Spergel, R. Bean, O. Dor , M. R. Nolta, C. L. Bennett, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, et al., astro-ph/0603449 (2006).